



Education
Endowment
Foundation

Lesson Study

Evaluation report and executive summary
November 2017

Independent evaluators:

Richard Murphy, Felix Weinhardt, Gill Wyness and Heather Rolfe





The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.



For more information about the EEF or this report please contact:

Danielle Mason

Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

p: 020 7802 1679

e: danielle.mason@eefoundation.org.uk

w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from the London School of Economics: Richard Murphy, Felix Weinhardt and Gill Wyness.

The process evaluation was carried out by Heather Rolfe from the National Institute of Economic and Social Research.

The lead evaluator was Richard Murphy.

Contact details:

Richard Murphy

Dept. of Economics, University of Texas at Austin

Tel:(001) 512 475 8525

Email: **richard.murphy@austin.utexas.edu**

Contents

Executive summary.....	4
Introduction	6
Methods	10
Impact evaluation	20
Process evaluation.....	34
Conclusion.....	47
References	50
Appendix A: Expression of Interest form.....	51
Appendix B: Memorandum of Understanding	53
Appendix C: EEF cost rating	57
Appendix D: Padlock rating	58

Executive summary

The project

Lesson Study is a popular approach to teacher professional development used widely in Japan. It involves a small group of teachers co-planning a series of lessons based on a shared learning goal for the pupils, with one teacher leading the co-constructed lesson and their colleagues invited to observe pupil learning in the lesson. The team then develop their practice further, based on feedback.

Some of the features of Lesson Study have been adopted, in a variety of formats, in the U.K. A team at Edge Hill University developed a Lesson Study programme combined with a Talk for Learning content focus, which involved a team of three teachers collaborating to plan, teach, (informally) observe and analyse the impact of teaching on pupil response and progress. Each teacher identified two 'case pupils' in their own class who appeared to be underachieving, and the observation focused on these pupils. The aim of the intervention was to raise pupil attainment by improving teacher effectiveness.

This evaluation tested the impact of a particular version of the Edge Hill programme on pupils' Key Stage 2 (KS2) outcomes. Participating schools nominated at least one group of three teachers for the programme. They attended five training sessions over the course of the intervention, which lasted from September 2013 to July 2015. Each term, they completed one Lesson Study cycle where one of the team taught three co-planned 'research' lessons while the others observed pupil learning, interviewed the case pupils after each research lesson, and met with the teacher to review their practice. Each year, at least one Lesson Study cycle focused on literacy, and at least one on maths; at least one cycle focused on a Year 4 class and one on a Year 5 class. Lesson Study is an approach that can work with a range of teaching content. To provide the content of the Lesson Study cycles, teams were trained in Talk for Learning approaches, which support children's learning through promoting positive talk in lessons.

The programme was evaluated using a randomised controlled trial of schools in the South West, East Midlands, and North West. Teachers in 181 schools (teaching 12,747 pupils) took part. The primary outcome was a combination of KS2 maths and reading test scores for pupils who were in Years 4 and 5 in the first year of the trial. The process evaluation involved observation of teacher training, analysis of surveys, school visits, and interviews with participating teachers and senior leaders.

Key conclusions

1. The project found no evidence that this version of Lesson Study improves maths and reading attainment at KS2.
2. There is evidence that some control schools implemented similar approaches to Lesson Study, such as teacher observation. This trial might, therefore, underestimate the impact of Lesson Study when introduced in schools with no similar activity. If that is the case, the results suggest that this version of Lesson Study had no impact over and above elements of the Lesson Study approach that were already widely used.
3. Teachers felt Lesson Study was useful professional development, valued the opportunity to collaborate with colleagues in a structured way, and reported several changes to their practice as a result of the programme.
4. Schools generally implemented the programme as the developers intended. Attendance at training was high and most schools implemented one Lesson Study cycle each term.

EEF security rating

These findings have very high security. This project was an effectiveness trial, testing whether the intervention was effective under everyday conditions in a large number of schools. The trial was a large, well-designed, two-armed randomised controlled trial. Relatively few pupils who started the trial were not included in the final analysis. The pupils in Lesson Study classes were similar to those in the comparison classes in terms of free school meal (FSM) entitlement, prior attainment, and gender balance.

Additional findings

Twelve months after the end of the Lesson Study programme, this evaluation found no impact of Lesson Study on KS2 reading and maths attainment, nor for any of the secondary outcomes (maths and reading individual scores; spelling, punctuation and grammar (SPAG) scores; and science scores), or on pupils with ethnic minority status, FSM eligibility, low KS1 attainment, or English as an additional language. There was no indication of different impacts for schools that delivered more lesson study cycles.

The process evaluation indicated that Lesson Study was implemented as intended in intervention schools. The programme was generally well received by both teachers and school leaders. Most schools were already implementing some form of lesson observation, but teachers valued the external training and the structured, timetabled peer collaboration that this programme offered; these were generally viewed as new and as supporting continued engagement with the content of teacher training. Teachers reported that their understanding of how to improve beneficial talk in the classroom was enriched through this programme.

Some teachers said they had changed their practice in relation to case pupils and reported improvements in these pupils' skills, for example in vocabulary, reading comprehension, and metacognition. Case pupils were not tracked individually so we do not have information about whether these pupils benefitted from the intervention.

There is limited information about control group activity, but there is evidence that some schools used elements of Lesson Study (such as teacher observation) or other programmes called Lesson Study, which were similar in aim to the Edge Hill programme but less structured (different packages are available for schools to buy and resources are available for schools to pursue independently.) This may have diluted the effect seen in intervention schools. The lack of impact in intervention schools suggests that the programme tested here does not have an impact over and above less formal and less structured elements of the Lesson Study approach that many schools already do.

Summary of impact on primary outcome—average Key Stage 2 test scores

Group	Effect size (95% confidence interval)	Estimated months' progress	No. of pupils	P value	EEF security rating	EEF cost rating
1 Year of Lesson Study vs. control	0.02 (-0.06 – 0.09)	0	6437	0.65	🔒🔒🔒🔒🔒	££££££
2 Year of Lesson Study vs. control	0.03 (-0.06 - 0.12)	0	6304	0.5	🔒🔒🔒🔒🔒	££££££

Cost

Lesson Study cost each school £2,300 per year (or £54 per student per year) when averaged over three years. Schools can now purchase Lesson Study training from Edge Hill University directly. Costs range from £1,800 to £2,400 per school. Costs can be reduced for groups of five or more schools.

Introduction

Intervention

Lesson Study is a process incorporating teacher observations and feedback. It is a professional development programme with a long history of use in Japan and is being increasingly used in the U.K. and worldwide. Teachers work in small groups to plan lessons that address a shared teaching and learning goal. For this intervention, these groups consisted of a learning tripod with three teachers from the same school, one of whom would be the selected on the basis of being the 'expert teacher'. Schools were free to choose which teachers would be involved in the intervention, with the restriction that two should be teaching year groups 4 and 5. Some of the smallest schools had mixed-age classes, and so one teacher may have taught both Year 4 and Year 5. Given the tripod design, if a school had only one class per year group, or less, they would have to choose a teacher from another year, which was seen as unproblematic from the Lesson Study developer's perspective. We placed no restriction on what other year group was chosen. Because the randomisation and analysis is at the school level, we are not concerned about schools being able to choose the teachers.

Schools used different criteria to select teachers to take part in Lesson Study. While all schools chose teachers with some subject expertise in English or maths as the expert teacher, selection of other members of the team sometimes included teachers whose practice was in need of improvement. In other cases, heads chose teachers who work well together and were therefore likely to make an effective Lesson Study team. In smaller schools with one teacher for each year group, there was little choice over the team composition and selection was not therefore strategic.

The leaders from the treated schools attended a senior leader conference held in London on 12 September 2013. This provided them with more details for how the programme had been implemented internationally and how it would be rolled out in their schools. This included discussion of the structure of the learning tripods, and the topics for the training days. The other purpose of this meeting was to generate enthusiasm in the leadership teams of the participating schools and included speakers from overseas.

In total, there were five training days for the teachers involved in the programme. This included two initial training days provided by Edge Hill University in September or October of 2013 to set up the learning cycle (details in Table 1.1 below). This was followed by a half day of training in November or December of 2013. At the end of the first year of the implementation (June 2014), there was a series of conferences in the different regions bringing teachers from each region together to share their experience of the programme for the day. At the start of the second year of the programme (October 2014), there was a training day focusing on optimising feedback and sustaining the programme through its second year. There was an additional sixth Continuing Professional Development day at the end of the programme (July 2015), which aimed to highlight the importance of evidence-based research in education and support Lesson Study teams to discuss their experiences and consider plans for sustaining Lesson Study.

Table 1.1: Lesson Study training

Introductory Conference (full day)

The potential of Research Lesson Study for School Improvement: the international context; the Edge Hill University Research Programme; leading and managing Lesson Study in school.

Day 1 (full day)

Part 1: Lesson Study Research Programme overview and the focus for researching pedagogy.

Part 2: Metacognition and Self-Regulation: developing and researching 'Talk for Learning' in English and mathematics.

Day 2 (full day)

Researching Pedagogy through Lesson Study: the Lesson Study process—ethos, protocols and practice.

Day 3 (half day)

Part 1: Reflections on first experiences and impact of Lesson Study.

Part 2: Taking the pedagogical focus deeper: researching 'Talk for Learning' through Lesson Study in guided sessions for literacy and mathematics.

Day 4 (full day)

Current Learning from the LS research programme and planning for sustainability: headteacher views; sharing findings from school visits; poster presentations by all Lesson Study teams.

Day 5 (half day)

Talking the pedagogical focus deeper: researching feedback on learning through Lesson Study.

The implementation of the Lesson Study cycle in schools started with an initial group meeting where the three teachers planned the order in which they were observed and which lessons the other teachers in the team observed. The first teacher then taught three 'research lessons' observed by the other two teachers who focused on pupil learning, rather than the teacher. After each of these lessons, there was a follow-up discussion that reviewed the class and joint planning for the next session in terms of content structure and delivery. In the planning, two 'case pupils' were selected that were typical of a group of pupils in the class so that the impact of the lessons on pupils with identified barriers to learning could be monitored. There was also input from the case pupil directly in the form of teacher–pupil interviews. Teachers were expected to keep records of the training sessions and were also expected to write up pupil case studies after each round of Lesson Study. This process was repeated after each of the research lessons. Over the course of the academic year, there were three cycles of Lesson Study with each teacher taking the turn of being observed. Fidelity with the programme implementation was measured through school visits and interviews.

The intention is that this structure provides a space for non-judgemental discussion in the school day and the use of case students helps focus the teachers. The trial of Lesson Study incorporated Talk for Learning, an approach which aims to improve the quality of classroom talk in order to increase pupils' engagement, learning, and attainment. Participating schools were asked to deliver Lesson Study within literacy and maths lessons using Talk for Maths and Talk for Literacy with teachers trained in specific approaches for both. The trial was therefore a test of the effectiveness of this method as well as of the Lesson Study approach.

To summarise, this trial of Lesson Study had a number of key features: (1) teacher peer-to-peer observation (of pupils' learning) and feedback, (2) teacher-pupil feedback, (3) Talk for Learning, (4) collaborative planning, and (5) case pupils. Some of these components are already used in English primary schools. Lesson Study provides structure through which teachers are able to gain more from each and achieve additional synchronicity.

Background evidence

It is widely accepted that teachers are the most important factor in school effectiveness (Rockoff 2004; Rivkin, Hanushek and Kain, 2005). However, there is little robust experimental evidence of the effectiveness of programmes that improve teacher quality. Of the high quality research that has been conducted on teacher development, knowledge based training has typically been found to be ineffective, whereas programmes that involve teacher observations that are embedded into the school alongside effective feedback have been found to have positive effects (Taylor and Tyler, 2012).

An evaluation of the U.K.'s National Strategies' Leading Teachers Programme (Hadfield, Jopling and Emira, 2011), which involved Lesson Study, showed that those schools using this approach (among others) out-performed a comparison group in both English and maths. Similar to this report, this involved a comparison of pupil progress in KS2 between treated and untreated schools within the same local authority (LA). They found that treatment pupils' KS2 results improved by 1.3% more than control schools in the same LA. However, it had three limitations: (1) it was not a randomised controlled trial (RCT), (2) it did not take account of potential differences between the schools, and (3) it was not able to follow up the cohort to explore any long-term impacts of the programme. Lesson Study also shares many of the key characteristics of effective CPD that were identified in a systematic review produced by the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI, 2003).

This EEF trial of the Lesson Study programme set out to robustly evaluate this intervention and is one of the largest randomised trials to date focusing on the effectiveness of this form of teacher development. Previous papers either used quasi-experimental methods (Taylor and Tyler, 2012) or were based on smaller samples (Lewis and Perry, 2015). Lewis and Perry implemented the only other RCT of Lesson Study involving 213 teachers (equivalent to 71 tripods) in 39 groups over an 80–91-day period. Using a three branch trial with control, Lesson Study, and Lesson Study plus additional maths materials, they found positive significant impacts on self-reported 'Collegial Learning Effectiveness' and 'Expectations for Student Achievement' when Lesson Study is combined with the maths materials. This group also saw improvements in the students' understanding of fractions, which was the focus of this Lesson Study trial. There were no improvements when Lesson Study was used in isolation. This report builds on their findings as this trial is closer to the realities of school life by being delivered over an extended period of time (two years) and looking at general student progress using KS2 data. Replicating similar conditions to this previous trial, this trial combines Lesson Study with Talk for Literacy and Talk for Maths to provide content for teachers to focus on in their lesson cycles. Talk for Literacy has been evaluated by the EEF previously with promising findings, although these were not quite statistically significant, and approaches that encourage talk in the classroom are increasingly popular in schools so fit with schools' existing foci. Finally, there has been no robust evaluation of this type of continuing professional development in the U.K.

Evaluation objectives

The central objective of the impact evaluation is to estimate the impact of the Lesson Study programme on pupil attainment, in particular on pupil KS2 results, which are the key metric of interest to schools and the Department for Education. Thus, the main outcome measure of interest is pupil academic outcomes as measured by KS2 tests. Specifically, our primary outcomes are KS2 combined average test scores across maths and reading for each cohort.

Ethical review

Before implementation, the research proposal was reviewed by the Ethical Review Board of the London School of Economics and Political Science to ensure no undue burden was placed upon the schools or students and individual anonymity would be ensured.

Schools were invited to participate in the trial, and those that were interested signed and returned an 'expression of interest' which states that they understand that the average student test scores from

schools involved in the trial would be obtained from administrative sources. The 89 selected schools signed an additional memorandum of understanding which states the responsibilities of the schools, practitioners, and the evaluation team.¹

Towards the end of the project, we also surveyed the headteachers—and other individuals responsible for continuing professional development—at all 181 schools regarding their training practices over the previous three years. This consisted of an emailed survey, pre-empted with a phone call to the school to give them notice that they should be expecting the email and so that it reached the most appropriate person. This element of the project underwent ethical review at the London School of Economics level.

Project team

Lesson Study was developed by a team at Edge Hill University with support from external consultants. The evaluation was conducted by Richard Murphy, Felix Weinhardt, and Gill Wyness for the London School of Economics, with the process evaluation led by Heather Rolfe for the National Institute of Economic and Social Research.

Trial registration

This trial was registered with AEA RCT Registry and assigned registry number AEARCTR-0001779.

¹ This project started before the current Department for Education (DfE) guidelines requiring parental consent. Regardless, the nature of the trial means that they are not required given Condition 6 of Schedule 2 of the DfE guidelines.

Methods

Trial design

This study entails a cluster randomised controlled trial evaluation of the Lesson Study programme blocked in the South West, East Midlands and North West of England. Schools were stratified by Local Authority and were pairwise randomly assigned to treatment or control arms. Further details of the sample and randomisation process can be found below. The trial lasted from August 2013 to June 2015 during which time teachers were trained in, and implemented, the Lesson Study programme. The year groups that these teachers instructed were Years 4 and 5. These students took their Key Stage 2 examinations in May 2015 and May 2016.

Randomisation was conducted at the school level, with the intention of having 80 treated schools and 80 control schools. Schools that were selected into the treatment group were asked to select three teachers to be part of the trial. Schools had complete autonomy over which teachers they chose, on the condition that these teachers were to teach Years 4 and 5. Where that was not possible, because there was only one class per cohort, for example, schools were free to involve teachers from other year groups. The involvement of teachers from other year groups is unproblematic, despite differences in curriculum between these years, as Lesson Study assumes that factors affecting student learning and in-class behaviour are similar across the year groups. This means that it is possible for students to be taught by Lesson Study teachers from Year 3 through to Year 6, but only students in Year 4 and Year 5 were included in the evaluation. Students who were originally in Year 5 were the first cohort of students to take their KS2 examinations and are referred to as 'cohort 1' (C1). The students who were originally in Year 4 were the second cohort of students to take KS2 examinations after the intervention and are referred to as 'C2' (see Table 2.1). Teachers of the target year-groups in treated schools were trained in the Lesson Study programme. The control condition was 'business as usual'; Lesson Study training and support was withheld from control group schools and they agreed not to pursue Lesson Study, although there are freely available materials online. Once control schools received the letter stating that as a result of the randomisation process they had not been selected, we had no further contact with them until after the trial. There was also no record made of what the control schools did, so we are reliant on the 11 completed post-trial surveys for fidelity information. However, we do have the student-level KS2 examination results for all control schools.

Table 2.1: Visualisation of treated cohorts within treated schools

Academic Year	2013/14	2014/15	2015/16	2016/17	2017/18
Year 2	C4				
Year 3	C3	C4			
Year 4	C2	C3	C4	..	..
Year 5	C1	C2	C3	C4	..
Year 6, KS2 exam	C0	C1	C2	C3	C4

The programme ran over two years (2013/2014 and 2014/2015) with the same schools being in the treated or control groups throughout the study. The last training day for the treated schools occurred in July 2015, although the second cohort of students would not take their Key Stage 2 examinations until May 2016 (see timeline in Table 3.1). As we highlight later on, this time lapse between programme and testing may have contributed to our inability to detect a significant effect of the programme.

Having a single programme allocated to a sufficient number of schools, as was the case here, has a number of advantages. First, applying the programme in isolation makes it possible for researchers to identify its effects. Second, randomisation into treatment and control groups ensures that the treatment and control schools have similar characteristics. Third, this school-level approach allows us to use administrative data taken from the National Pupil Database, giving us sufficient sample sizes necessary to make proper inference, and a long term series of results by school, which can be further exploited by the evaluation strategy.

Our evaluation strategy entails the use of (1) a basic differences method—comparing KS2 outcomes of the treatment and control groups (obtained from the National Pupil Database), and (2) difference-in-difference analysis—comparing changes over time in the results of treatment schools with those of control schools.

A further advantage of this strategy is that as the compulsory test scores are collected centrally we did not need to be concerned with attrition due to testing in treatment or controls schools. This will ensure internal validity of the results.

Participant selection

The target population for this study is state primary schools in England with an above average proportion of FSM-eligible pupils (which stood at 19% in 2013: National Statistics, 2016) and two or fewer classes per cohort. Note again that since the Lesson Study approach does not propose to develop teaching skills specific to a particular year group, the involvement of teachers from different year groups is unproblematic.

The project developers were asked to approach such primary schools in three regions to recruit into the study. The regions were the South West, East Midlands, and North West. A regional approach was chosen to minimise travel time for trainers and participants relating to training sessions. Within these schools, the target teachers were those teaching in academic Years 4 and 5.

The aim of the recruitment was to eventually have 160 schools participate in the study. This total was determined by baseline power calculations (full details can be found in the sample size calculation below).

Of the 182 schools recruited to this trial, one was ineligible as it would not have a Year 6 during the years that the target cohorts would be taking their KS2 examinations and so was excluded. This left 181 schools that were all eventually randomised to treatment or control status as described in the randomisation section below. All of these 181 schools returned a signed expression of interest (Appendix A) which stated that their school data would be used via access to the National Pupil Database (NPD). The 89 selected schools signed an additional memorandum of understanding (Appendix B) which stated the responsibilities of the schools, practitioners, and the evaluation team. This culminated in a total of 6,441 students who formed the first cohort of students, and 6,306 in the second cohort.

Outcomes measures

Primary outcome

The main outcome measure of interest is pupil academic outcomes as measured by Key Stage 2 tests. Specifically, our primary outcome is KS2 combined average test scores across maths and reading for each cohort. These test scores were obtained via the NPD, which Murphy, Weinhardt, and Wyness

applied for through the usual official channels. This database includes information on KS2 test results (for reading and maths) as well as KS1 test results for basic reading, writing and maths. We primarily focused on KS2 test scores as the measure of pupil achievement for all pupils. Due to the change in KS2 assessment, the KS2 composite English score was not used as low correlations between reading and writing have recently been reported (Allen, 2016).

All KS2 scores were converted into percentage scores at the national subject-cohort level to achieve comparability across subjects and years/tests. This ensures that there is comparability in the distribution of test scores across schools and years. This is important given the changes that occurred to KS2 assessment during this period. The exception to this is science, which is only reported in levels and so it is not appropriate to be converted into percentages. Moreover, there is no science outcome for the second cohort as it was not recorded in 2015/2016.

Secondary outcomes

Our pre-specified secondary outcomes are:

- KS2 maths test scores;
- KS2 reading test scores;
- teacher-assessed KS2 science level (cohort 1 only); and
- KS2 SPAG test scores.

Again, the maths, reading, and SPAG test scores were converted into percentages to account for any change in the test score distributions over time.

The rationale for focusing on KS2 outcomes is:

- These were the first national standardised tests the pupils took following exposure to Lesson Study and we were confident that this outcome variable would be sufficient to measure the effectiveness of the programme.
- KS2 data is the measure used to assess pupils at the end of primary school and we therefore wanted to explore any impact of the programme on this outcome measure.
- As these are national tests, this allows us to use administrative data taken from the NPD giving us sufficient sample sizes necessary to make proper inference, and a long time series of results by school which were further exploited in the evaluation strategy (and indeed could be tracked beyond the period of evaluation).
- These compulsory test scores are collected centrally so we did not need to be concerned with attrition due to testing in treatment or controls schools. This also ensured internal validity of the results.

Specifically, these outcomes capture pupil academic ability at the end of primary school when the pupil is age 10–11. Pupils are tested in maths, reading, and writing.² As these exams are compulsory national exams, they are administered internally by schools, but externally moderated by qualified teachers.

Sample size

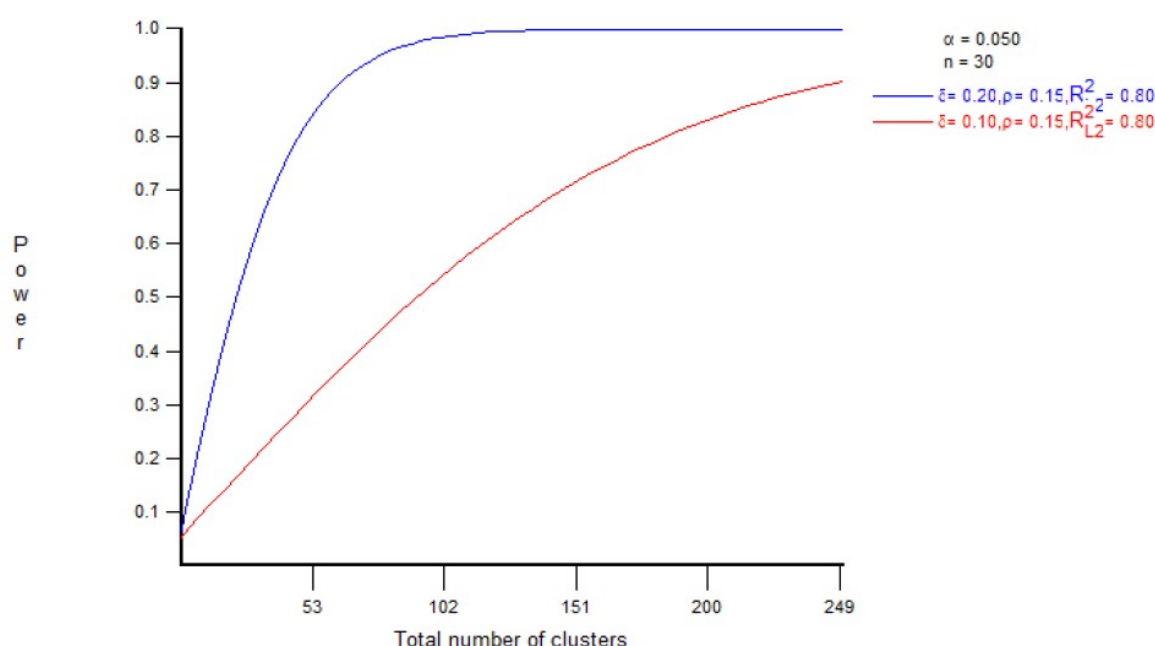
Assuming an intraclass correlation coefficient of $\rho = 0.15$, covariates including prior test scores explaining 80% of the variation in test scores, a class size of 30, and requiring a 5% significance level, power analysis revealed that a large number of clusters (in our case schools) would be needed for having a good chance to detect effect sizes of 0.1–0.2 standard deviations. This was using the Optimal Design software (Raudenbush *et al.*, 2011). As a result, the Lesson Study trial was restricted to two

² See <https://www.gov.UK./national-curriculum/overview> for more details.

arms (treatment and control) and all efforts were undertaken to include as many schools as possible, subject to budgetary constraints. The resulting sample of 182 schools (not controlling for dropouts or exclusions) lies in a power region where the detection of true effect sizes of 0.2 of a standard deviation in the outcome is very likely. However, an effect of this size would be considered large in the literature, and this study only detects true effect sizes of 0.108 SD in the outcome with a probability of about 80% (79.4% with 182).

Because we could not identify students to teachers, the ideal was to have only one class per cohort. However, in reality many schools had two classes per year group. We used a cohort size of 30 to be conservative. Re-calculating the power calculations with 60 students within a school instead of 30, with the same set of assumptions, means we reach a power of 0.8 at 144 schools, instead of 0.1 with 184 schools.

Figure 1: Power curves by number of primary school clusters (using Optimal Design software, Raudenbush *et al.*, 2011)



Randomisation

Lesson Study was able to recruit 182 schools across the three regions (South West, East Midlands, and North West). One of these schools was ineligible as it would not have a Year 6 during the years that the target cohorts would be taking their KS2 examinations and so was excluded. This left 181 schools that were all eventually randomised to treatment or control status. Richard Murphy performed a stratified randomisation of schools by Local Authority with the aim of balancing the randomisation at LA level, for example, the pairing of schools for randomisation was conducted within each LA. This was to ensure there were equal numbers of treated and control schools within each region and that they would be balanced in terms of unobservable regional characteristics.

We first generated a single index intended to match similar schools, among the volunteer schools, on the basis of multiple characteristics. This comprised students' average KS2 outcomes in English and maths, and the proportion of FSM students who took their KS2 tests in the summer of 2011. To create a single index, mean values for each school in 2011 were combined using Principal Component Analysis (PCA). There were eight schools that did not exist in their current form in 2011 (five due to not having a Year 6, and four being new as a result of an amalgamation) and so had no KS2 or FSM-

eligibility information from which to create an index; these were, therefore, excluded from this index-generating process. In addition, before the first randomisation—to avoid the potential complicating issue of school-clusters—all clustered schools (six—three pairs) were removed from the first randomisation pool, along with those with no index, 14 schools in all. These schools would be used in reserve if many schools from the first randomisation no longer wanted to continue with the trial, and would be randomly allocated to treatment or control in the same way as the original randomisation.

The remaining 167 schools were then sorted according to their LA and then by their index score, ensuring that similar schools were ordered next to each other. Pairs were then formed by taking the first two schools within a region, and then each subsequent pair. All remaining un-paired schools from a LA with an uneven number of schools were then put into another pool and ordered according to the index. Again, schools were paired according to their next school. One school was left un-matched. This process caused slight imbalances in the number of treated and control schools within each LA.

A random seed was set and then each school was then given a random number from a uniform distribution. The school with the higher random number within each pair was then chosen as the treatment school; 83 schools were assigned to treatment and 84 were assigned to control. The one unmatched school was assigned to a control status. Manually assigning any unmatched school to the control group was part of the research design. This is because unmatched schools can have no effect on any of the results as any un-matched schools cannot contribute to the main analysis as pair fixed effect would absorb the average test scores of students enrolled at this school. Therefore, it was decided to assign this school to the control group and to maximise the limited budget by not carrying out the treatment in this school.

These 83 schools were invited to take part in the treatment; 16 of these randomly-allocated schools subsequently declined to take part. We surveyed the reasons for this early dropping-out: none seemed related to the content of the Lesson Study intervention itself—nine of these 16 schools cited staffing changes as the main reason. This freed up some space for all the remaining schools not included in the first randomisation to be used in a second randomisation (to treatment or to control). The second randomisation used a similar method to the first: pairing schools, and then allocating the school with the higher random number (within the pair) to the treated group. This pairing occurred within groups relating to the reason for exclusion from the first randomisation to ensure similarities. For example, all the schools that were excluded for not having prior KS2 test scores were put into a single grouping and assigned randomised numbers; alternate schools were then allocated to the treatment group, starting with the highest number. A similar process was followed with cluster schools and amalgamations. From this randomisation, six were assigned to treatment and nine to control. This imbalance is due to an odd number of schools within the cluster grouping and the new Year 6 grouping, with the unmatched schools set to be in the control group. To be specific, there were three school clusters; one cluster was left unmatched (two schools). In addition to this, there was an uneven number of schools that had a new Year 6: this left another school unmatched. This left a total of three unmatched schools from the second randomisation.

Therefore in total 89 school were allocated to treatment: 83 from first randomisation (16 of which declined to take part) and 6 from second randomisation. For the control group, 92 were allocated: 84 from first randomisation and 8 from second randomisation), four of which are unpaired. We will return to these issues of school selection when describing Figure 2 and the impact assessment below.

Analysis

Analysis was conducted at the pupil level, clustering at school level.

Our analysis plan in the protocol was different from the current EEF guidance; therefore, both sets of estimates (primary analysis and protocol analysis) were produced as described below. All econometric analysis was conducted with the Stata software package.

Primary Analysis

In accordance with the EEF guidance, the analysis of primary and secondary outcome measure(s) was undertaken on an intention-to-treat (ITT) basis meaning that all those allocated to treatment and control in the randomisation are included, even if they dropped out of the treatment. This means that for all analyses the maximum 182 schools was used. It should be noted that the schools that were left unmatched did not contribute to the estimation of the effect size in specifications with pair fixed effects as their average test scores would have been absorbed by this effect, which in these cases only represents one school.

The effect size on the primary outcome was calculated using OLS at the pupil level for increased power and to reduce bias, with clustering at the school level and reporting robust standard errors. Although a direct comparison of the means should be sufficient for determining the effect size, controlling for prior attainment (KS1) increases the precision and reduces any potential biases. Adding covariates reduces the total variances to be explained. For comparability, additional covariates were not included, with the exception of the pair-fixed effects. The estimated model therefore includes treatment status, KS1 attainment, and pairing as fixed effects using robust standard error. This model was estimated separately for cohort 1 (C1) and cohort 2 (C2). These effects are estimated separately so that no assumptions were made regarding dosage, for example, combining them in one specification and introducing a dosage term would assume linearity in dosage.

When determining the effect size, we use the total variance rather than the residual variance from the clustered model. Variations in a post-test outcome are due to different sources, which must be fully accounted for in a statistical model. For cluster randomised trials, the total variability can be decomposed into random variation between pupils (σ_i) and heterogeneity between schools (σ_s). Effect sizes for a cluster randomised trial with equal cluster size and using total variance are calculated as:

$$Effect\ Size = \frac{(\bar{Y}_T - \bar{Y}_C)}{\sqrt{\sigma_i^2 + \sigma_s^2}} = \frac{\beta_{Treat}}{\sqrt{\sigma_i^2 + \sigma_s^2}}$$

We calculated the effect of the programme on the first cohort (C1) and the second cohort (C2) for the average KS attainment. An equivalent table for the secondary outcomes has also been produced.

Protocol analysis

Here, we estimate the impact of the programme on pupil outcomes using a difference-in-differences approach using the following model.

$$Y_{ist} = \alpha + \beta(Treat_s * Cohort_t) + \lambda X_i + \theta_t + \mu_s + v_{ist} \quad (1)$$

The dependent variable Y is the pupil i KS2 test score from school s in year t . β represents the effect of the programme on pupils; $Treat_i$ and $Cohort_i$ are indicator variables which will equal 1 for treated schools and students who took their KS2 examinations in the academic year corresponding to the appropriate cohort (C1 = 2014/2015, C2 = 2015/2016), and 0 in all other circumstances. We also include a vector of pupil characteristics X_{is} to take into account their effects on test scores, which improves the efficiency of the estimations. The student characteristics are KS1 test scores in the relevant subject (previously included), gender, and FSM status when the student took KS1. Finally, we include a set of school μ_s and year effects (θ_t) to control for any unobserved differences between schools, or across years; this further improves the efficiency, increasing the likelihood of estimating significant results. We cluster the standard errors at the unit of treatment, which is the school (school cluster).

We ran this model separately for each cohort. To avoid the potential issue of students in the pre-treatment years obtaining treatment (for example, the school provides the treatment to Year 6 pupils), the difference-in-difference analysis includes all years from 2008/2009 up to the start of the trial (2012/2013) as control years and the treatment year only. This means for C1 we omit 2013/2014 from the analysis and we omit 2013/2014 and 2014/2015 for the analysis of C2.

Secondary analysis

A key concern is that of non-compliance of treatment and control schools (never-takers and always-takers). Equation 1 represents the ideal situation with complete compliance and would provide an unbiased estimate of the Average Treatment Effect (ATE). In reality, some schools that were asked to be in the treatment did not accept (see above) and therefore we replace the treatment variable with one for intention to treat, $Intend_i$. Estimating this equation will give us an unbiased ITT effect. We define 'compliance' of treated schools as our having received confirmation from the school at the end of each academic year that they participated, and having received a UPN list of the students taught by a Lesson Study teacher.

However, policy makers are typically interested in the ATE, that is, how big the effect would be on an average school, or student. To address this issue we use an Instrumental Variable (IV) approach. This uses the original assignment lists to predict whether the school (student) will be treated. For this, $Treat_i$ indicates whether the school actually went through with the programme, and $Intend_i$ is an indicator variable for if a school was randomly assigned to be treated. $Intend_i$ will be a strong predictor of whether or not the school was actually treated, but as it is randomly assigned we know that it is independent of pupil outcomes and therefore using this as an instrument will account for bias due to non-compliance.

$$Y_{ist} = \alpha + \beta(\overline{Treat}_s * Cohort_t) + \lambda X_i + \theta_t + \mu_s + \varepsilon_{ist} \quad (2)$$

$$\overline{Treat}_{st} = \alpha + \beta_0(Intend_s * Cohort_t) + \lambda X_{is} + \theta_t + \mu_s + \vartheta_{ist}$$

We also produce an ATE using variation at the student level. Note that although an entire school is allocated to treatment or control, not all students within a school year are necessarily treated. This could occur for two reasons. First, Lesson Study has learning tripods (involving three teachers); if a school had two classes per cohort and they are split over Years 4 and 5, then one teacher in one of the treated cohorts would not be involved in the programme. The NPD data does not allow us to determine how many teachers are in a school year, but there is indicative evidence that this is the case. The proportion of a cohort being treated only fell below 50% (in treated schools that participated in the study) when the cohort size was above 34. Second, some students joined the school during Year 6, meaning they take the Key Stage 2 examinations with the treated cohort, but were not necessarily exposed to a lesson study teacher. Therefore, we obtained the UPN of each student taught by a Lesson Study teacher, and so we know whether or not pupils were exposed to the programme. For this student-level analysis, $Treat$ will indicate whether the pupil actually went through with the programme and $Intend_i$ is an indicator variable for whether a school was assigned to be treated. Again, $Intend_i$ will be a strong predictor of whether or not the pupil was actually treated, but as it is randomly assigned we know that it is independent of pupil outcomes.

$$Y_{ist} = \alpha + \beta(Treat_i * Cohort_t) + \lambda X_i + \theta_t + \mu_s + \varepsilon_{ist} \quad (3)$$

$$Treat_i = \alpha + \beta(Intend_s * Cohort_t) + \lambda X_{is} + \theta_t + \mu_s + \vartheta_{ist}$$

The results for these analysis will presented in the Impact Evaluation section below.

Missing data

As we relied on NPD tests, there should not have been any missing data. Those individuals who could not be linked by the DfE were dropped from the analysis. Very few students were linked but did not have measures for KS2 test scores: for the primary outcome measure, 13 treatment and 0 control pupils in the first cohort had no measure, and 1 treatment and 5 control pupils in the second cohort had no measure. For the secondary outcomes, the numbers were slightly higher, reaching 48 treatment and 31 control pupils for cohort 2. These observations are noted in the results tables. Due to the few

incidences, no imputation was carried out regarding missing data (the number of missing KS2 test scores can be seen in Tables 5.1 and 5.2; the maximum proportion missing is 1.2% for C2 Maths). Some schools in the analysis did not have a Year 6 in 2010/2011 when baseline balancing was calculated due to being a new school or as a result of an amalgamation. For these, no baseline school-level information is available and therefore they were included in the second round of randomisation. One school that volunteered for the programme had had no Year 6 by the time of the evaluation of the cohorts and so was excluded from the sample and was left unassigned. Overall, due to the reliance on the national Key Stage 2 exams, we have outcome data available for 100% of the schools in our treatment and control groups.

On-treatment analysis

The analysis at the school, cohort, and individual level was estimated for each cohort separately.

Secondary outcome analyses

The secondary outcomes were assessed using the same methodology and approach as the primary outcomes.

Subgroup analyses

The NPD data allowed us to measure outcomes for different types of pupil, and hence to detect heterogeneous impacts. We carried out the analysis on subgroups relating to (1) Free School Meal eligibility, (2) English as an Additional Language, (3) low achievers (as defined by KS1), and (4) ethnicity. We specified in the protocol that we would also carry out the analysis according to the Lesson Study specialisation of the school (numeracy or literacy), however schools were encouraged to have a dual focus on both subjects, with at least one cycle of lessons focusing on each in each academic year, so this was not carried out.

The primary subgroup analysis produced intention-to-treat estimates using the initial allocation to treatment and control groups. The size of the differences of the effects between the groups was determined by an interaction term of assignment to treatment and the characteristics of interest, while also controlling for that characteristic. The secondary subgroup analysis estimated the ATE at the student level (Specification 3) because the subgroups are at the individual level. Again, there were two endogenous parameters (treatment and treatment*characteristic) and two instruments (intend to treat, intend to treat*characteristics), and the difference between the groups was determined by the interaction term.

Implementation and process evaluation

As well as the quantitative component of the study described above, a process evaluation was also carried out. The principal aims of this were to understand teachers' and senior leaders' perceptions of the programme and identify factors affecting its successful implementation.

The process evaluation consisted of the following components.

- Evaluation of teacher preparation and training through:
 - attendance at the introductory conference in September 2013;
 - observation of teacher training in one regional centre in October 2013 and follow-up training in June 2014;
 - analysis of training evaluation surveys collected by the internal evaluation team at Edge Hill University; and
 - analysis of data on control schools' use of peer observation and Talk for Learning approaches.

- Evaluation of implementation, fidelity, and impact through:
 - visits to ten schools in two of the three implementation regions in March 2014 to interview 19 staff and senior managers involved in the implementation of Lesson Study as team members or expert teachers;
 - follow-up interviews by telephone and email in November 2014 with five expert Lesson Study teachers in five schools, with information on progress provided by four other schools; and
 - analysis of school visit data collected by the internal evaluation team at Edge Hill University.

The process evaluation was carried out by Heather Rolfe (NIESR).

Lesson Study was implemented in three regions of the U.K.: the North West; the South West, and East of England. The Edge Hill University research department covered all schools and conducted more in-depth research in a sub-sample of the schools. To avoid excessive demands on schools, the case study research was confined to schools in the South West and East of England. The Edge Hill University team carried out internal evaluation research in schools in the North West, which was used as additional information by the external evaluation. Our visits and follow-up interviews enabled us to gain a more in-depth understanding of the implementation of Lesson Study, while our own assessment of the project team data from visits and evaluation of training (based on raw data rather than interpretive reports) offered a broader understanding of teachers' experiences across all schools involved in the project.

Data, collected through the range of approaches described above, was analysed using a framework approach. This enabled the analysis of process data in a written form, and is therefore appropriate for the analysis of transcripts of interviews with teachers and senior leaders, as well as research notes taken during observation of training. Qualitative responses to training evaluation surveys were also analysed in this way. The method entails coding the data into themes and issues. In this case, codes were a mixture of predetermined themed ones, developed during the design of the process evaluation taking account of the aims of the intervention, as well as those that emerged from the text of transcripts and observations. Codes identified different types of information, for example, more tangible ones—such as previous use of Lesson Study and Talk for Learning approaches, experiences of the training, and of putting the approach into practice—as well as others such as values and feelings. Throughout the analysis process, we looked for similarities and differences in the data. The framework approach allows for tracts of text to be classified under more than one code, and codes were, in some cases, amalgamated to form wider groups, particularly where substantial issues were concerned.

The codes and groups developed in the analysis of data formed the analytical framework and were used to structure the findings into a preliminary report. The framework was formed from the case study visits data and was then used to structure findings from the follow-up interviews, project team visits, and training observations and evaluation. In the report, where relevant, we identify the source of a finding, for example, whether from the initial case study visits, or project team visit.

Costs

The costing information was collected from Edge Hill University at the end of the trial. This costing accounted for the total implementation of the programme excluding any evaluation costs (a full breakdown of costs can be found in Appendix A). The cost per school per year was calculated on the basis on the number of schools that ultimately participated (74) and as if they had participated over three years (per EEF advice) assuming that the programme would have no financial cost in the third year. The cost per student per year is estimated according to the number of students in the treated school cohorts as if the programme had been implemented for three years. This is assuming all students in a school year group were treated, which may be an over-estimate (or more likely an under-estimate) as the Lesson Study model requires three teachers and there are typically fewer than three classes per

year group in English schools. We also provide costings for the programme as currently provided by Edge Hill.

Timeline

Table 3.1: Timeline

Date	Activity
May–June 2013	Approach schools to gauge interest in participation in Lesson Study
June 2013	Edge Hill supply list of interested schools
June 2013	First randomisation of schools into treatment and control groups
June 2013	Recruitment of selected schools into the programme
July 2013	Second randomisation of schools into treatment and control groups
July 2013	Recruitment of selected schools into the programme
September 2013	Introductory conference
September 2013	Implement Lesson Study in selected schools
March 2014	Process evaluation of implementation, fidelity in 10 schools
November 2014	Interview with five Expert Lesson Study Teachers in five schools, and progress in four other schools
May 2015	Cohort 1 take Key Stage 2 tests
July 2015	Programme officially ends
May 2016	Cohort 1 take Key Stage 2 tests
June–July 2016	Post-trial survey of treatment and control schools
Feb–March 2017	Analysis
March 2017	Report writing

Impact evaluation

Participants

A total of 182 schools were recruited for the Lesson Study evaluation. A particular feature of this intervention is that the research team did not collect any student-level information, for example for baseline comparisons. Instead, this analysis relies entirely on linked-in administrative data that is available for all students in state education in England. This particular approach was taken to maximize the number of schools participating while minimizing costs. Had the team conducted student-level testing itself, the funding allocated would not have allowed such a large trial. This particular approach has important implications to be aware of when considering participation. As a direct result of this design, we do not know the initial number of students treated by the intervention with absolute certainty as we rely on schools accurately reporting the UPNs of the students in the classes of the Lesson Study teachers. Instead, using the administrative data from the National Pupil Database, we do have information on all students in the academic years and schools where treatment was assigned. This approach has the additional advantage that there is minimal attrition due to testing of all students. By relying on the pre-treatment KS1 test scores as well as the nationally representative KS2 tests we ensure having information on virtually every student in treatment and control schools.

Figure 2 shows the participant flow from the recruitment to the analysis stage. Since recruitment, randomisation, and analysis is mainly conducted at the school level, this figure displays the changes in the number of schools throughout the process. In the recruitment stage, an unknown number of schools was approached and asked to express interest in participating in our trial. This either happened through talking to schools and teachers directly or through the Local Authorities. To signal such an interest, schools had to sign and send back an expression of interest (Appendix A). In a first step, we excluded 14 schools not meeting our predefined eligibility criteria from first randomisation, and randomised over two stages to assign a total of 92 control and 89 treatment schools. Details of the randomisation are explained in the section on randomisation above.

Of the 89 treatment schools, 16 declined treatment at the point of being asked to sign a memorandum of understanding (Appendix B) at the beginning of the intervention period, five dropped out during the first year, and another four during the second year of the trial. As a result, 64 of the 89 schools assigned to treatment actually went through the whole two-year intervention. Since dropout can be non-random, the following descriptive analysis of the baseline data as well as the main results section are based on school assignment rather than realised participation. However, in the analysis section, we also present average treatment effect (ATE) results for those schools and pupils who were actually treated.

Figure 2: Participant flow diagram

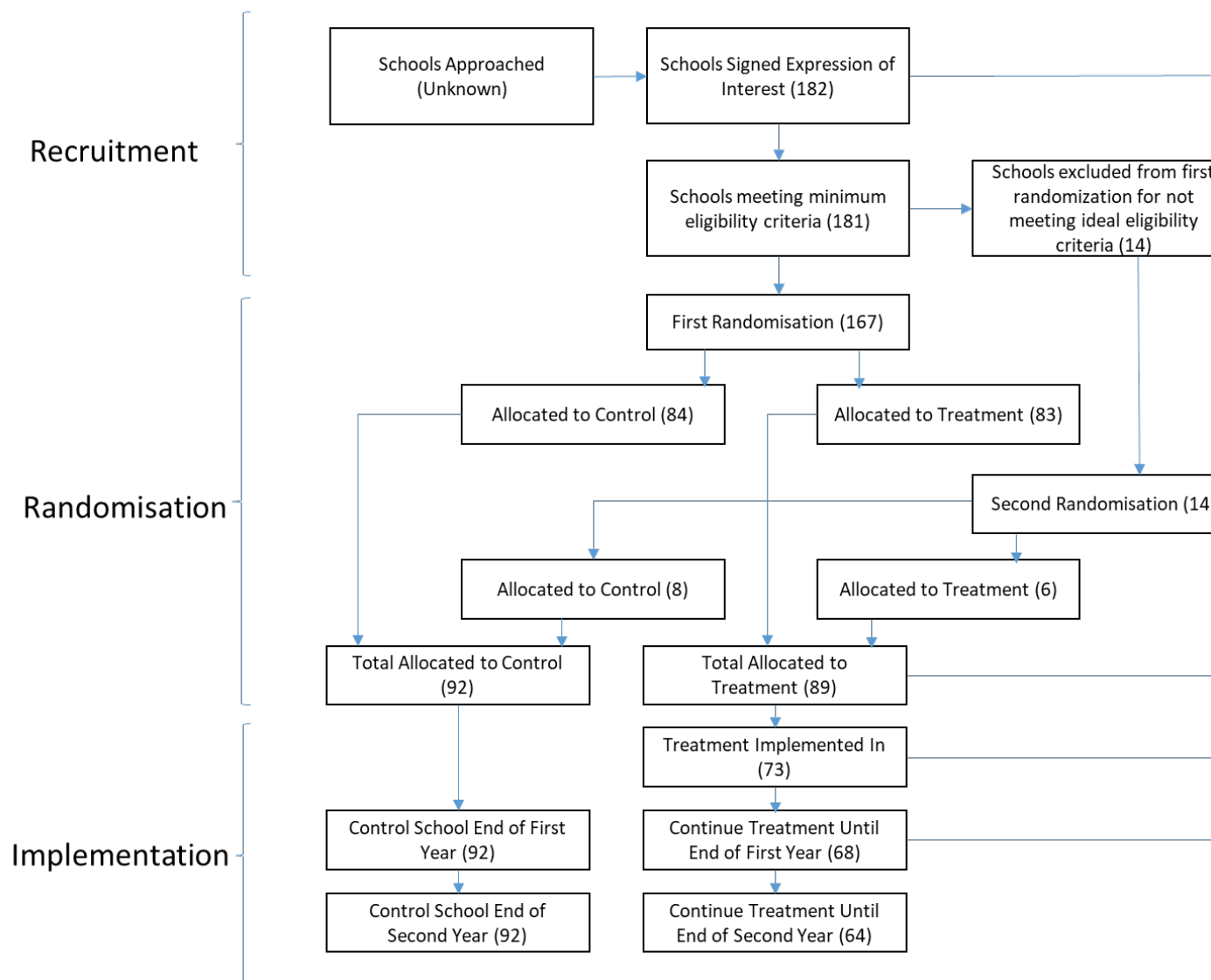


Table 3.2: Minimum detectable effect size at different stages

Stage	Schools (Pupils) (intervention; control)	Correlation between KS1 and KS2 percentile	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	160 (30) (80; 80)	0.8	0.15	Pair Matching	80%	0.05	0.108
Randomisation	181 (46) (89; 92)	0.717	0.115	Pair Matching	80%	0.05	0.095
Analysis (i.e. available pre- and post-test)	181 (46) (89; 92)	0.717	0.115	Pair Matching	80%	0.05	0.095

The experiment was designed to have a minimum detectable effect size of 0.1 (0.108), assuming 160 schools and 30 students per school cohort. In reality, we had 181 schools with an average of 46 students per cohort. Despite prior test scores having a lower correlation with outcomes than predicted the minimum detectable effect size remains high at 0.095 (see table 3.2).

Pupil characteristics

Table 4.1 shows summary baseline statistics for the treatment and control schools, both at the school and at the student level.

The first row, which shows school-level averages of FSM eligibility across treatment and control schools, clearly documents that the intervention was successful in targeting disadvantaged schools. In both groups, a quarter of students were FSM-eligible in 2011 (as defined by the standard NPD release definition of 'Is student known to be eligible for FSM?'), which is much higher than the national average for this year (18%).

The second panel, which shows continuous variables for treatment and control schools, reinforces this message: students in both treatment and control schools, on average, underperformed in KS2 tests in 2011 in mathematics and reading. For comparability, these test scores are converted into percentiles with a national average of 50. All four cells (maths, reading, in control and treatment group) show average student attainments well below 50, with a particular weakness in reading. The third and fourth panel, which shows student-level data separately for treatment and control schools, confirms these findings.

The second key take-away from the baseline comparison is that there are no significant differences in these characteristics between the treatment and control groups. This effectively confirms that the randomisation was effective and equivalence was observed. Note that this is a particularly strong result for this study since we can show that treatment and control schools are balanced not only on baseline characteristics such as Key Stage 1 tests, but also on the actual outcome of interest (Key Stage 2 tests) from an earlier cohort. Usually, the outcome of interest is not available for pre-treatment cohorts but we can show these results precisely because we rely on administrative information rather than tests conducted by the evaluators directly.

Table 4.1: Baseline comparison, intent to treat sample

Variable	Intervention group		Control group	
School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for free school meals	89 (0)	26%	92 (1)	25%
School-level (continuous)	n (missing)	Mean	n (missing)	[Mean]
School Size	89 (0)	265.8	89 (4)	266.74
Average KS1 Maths (2011)	89 (0)	15.42	89 (4)	15.31
Average KS1 Reading (2011)	89 (0)	15.32	89 (4)	15.23
Average KS2 Maths (2011)	84 (5)	46.28	85 (8)	47.92
Average KS2 Reading (2011)	84 (5)	45.03	85 (8)	46.03
Pupil-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
Eligible for free school meals	3,230 (0)	24%	3,329 (0)	24%
Male	3,230 (0)	50%	3,329 (0)	51%
Pupil-level (continuous)	n (missing)	Mean	n (missing)	[Mean or median]
KS1 Maths test score	3,227 (3)	15.50	3,328 (1)	15.38
KS1 English test score	3,229 (1)	15.43	3,328 (1)	15.36

As we rely on NPD tests, there is little missing data. If some individuals for whatever reason could not be linked by the DfE to the NPD, the team refrained from imputation and these individuals were excluded from the analysis. This was the case for only a very small number of students as indicated in the brackets above.

However, there were some schools in the analysis that did not have a Year 6 at baseline, that is, before the intervention, and so had not tested students (or did not exist in 2010/2011 because of being a new school as a result of an amalgamation). For these schools, no baseline school-level information is available and therefore they were counted as missing and not included in the balance tables. Following our protocol, we refrained from imputing any baseline information on these schools.

Outcomes and analysis

Section 5: Estimates

In the following we provide the main results of the trial analysis. We do this in three sections.

We start by reporting post-treatment differences between treated and control cohorts. In all tables we show descriptive statistics for treatment and control groups separately, as well as reporting Hedges g effect sizes when controlling for prior test scores and pair fixed effects (with 95% confidence intervals) and their associated p-values. This is in line with updated EEF analysis guidelines and is described in the analysis section as 'primary analysis'. The results will be discussed in Section 5.1.

The second section (5.2) then presents coefficient estimates of OLS regressions using the ITT framework, where all students in treatment and control schools are used in the analysis regardless of whether they ended up receiving treatment. This is in line with what was outlined in the protocol and is described in the analysis section as ‘protocol analysis’.

The third section (5.3) follows the IV approach as set out above in order to estimate treatment effects for only schools and pupils who were treated. In contrast to the comparison of means and the ITT, the IV analysis estimates effects only using schools and pupils actually involved in the lesson study treatment. This analysis corresponds to equations (2) to (5) above.

Finally, Section 5.4 presents subgroup analysis for students with FSM status, EAL status, minority status, or low KS1 score.

In all cases, we present results for our primary outcome of average KS2 results. In addition, we report separately results for our secondary outcomes—KS2 maths and reading, SPAG, and science scores.

Section 5.1

Table 5.1 reports the raw means of our primary outcome for treatment and control groups and both cohorts separately. The totals represent the number of students that we have valid KS2 test scores for the intervention and control groups respectively. The value in parenthesis below represent the number of student for which we do not have a valid KS2 score and is considered missing, this is typically less than 1%. In principle, cohort 1 received treatment over two academic years whereas cohort 2 was only treated in one academic year, assuming that schools did not extend Lesson Study into the final year of primary education (Year 6) after the actual intervention had faded out. In principle, this would also allow estimating effects of different dosages. However, Table 5.1 shows no significant differences in raw means between treatment and control groups, independent of dosage. The KS2 average percentile for the first treated cohort was 47.23, for the control group it was 46.11. While the intervention group of cohort 1 obtained slightly higher average KS2 results compared to the control group, these differences are negligible. The computation of the effect sizes confirms this assessment; this additionally controls for prior KS1 attainment and pair fixed effects. Note that for some students there is no valid KS1 test score and so cannot be included in the model. For cohort 1, this applies to three students in the intervention group and one in the control group. These new totals can be seen in the third to final column. The estimated Hedges g effect size is 0.02 and a 95% confidence interval that comfortably sits on the zero threshold. The corresponding p-value confirms that this effect is not statistically significant at conventional levels.

The second panel of Table 5.1 shows identical statistics for the second treatment cohort. Here, the average score of the treatment (control) group is 45.49 (45.49) points, and there is thus again no significant over- or under-performance of the treatment group.

Table 5.1: Primary results

Outcome	Raw means		Raw means		Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS2 Average							
Cohort 1	3,157	47.23	3,284	46.11	6,437	0.02	0.65
	(13)	(46.34–48.12)	(0)	(45.25–46.98)	(3,154–3,283)	(-0.06–0.09)	
Cohort 2	3,177	45.49	3,129	45.49	6,304	0.03	0.5

	(9)	(44.61– 46.36)	(5)	(44.60– 46.39)	(3,177– 3,127)	(-0.06– 0.12)	
--	-----	--------------------	-----	--------------------	-------------------	-------------------	--

We now turn our attention to the secondary outcomes, namely KS2 scores in maths, reading, SPAG, and science. Again, we present results separately for cohorts 1 and 2, with the hypothesis that effect sizes should be larger for the first cohort. Note that because the model effect size is additionally conditional on KS1 test scores and pair fixed effects, it is possible for the sign of the effect to be different to the raw differences.

The raw mean of KS2 maths percentiles of the first cohort is 48.1 (46.24 in control group). Again, this difference is positive, which means that the students of the treatment group do outperform students of the control group. However, this difference is again very small in magnitude and the Hedges g effect size calculation confirms that this small difference is also not significant at conventional levels.

In the second cohort, which was less affected by the treatment, the differences in KS2 maths outcomes between treatment and control groups are even smaller. The treatment group averages at 46.49 percentile points whereas the control group students obtained an average of 46.09 percentile points. Again, there is a positive difference, but this difference is even smaller and again not significant at conventional levels.

Next, we turn to KS2 reading outcomes. From our balancing tables above we know that the students and schools participating in the Lesson Study trial were particularly underperforming in KS2 reading scores before the intervention in 2011. Unfortunately, there is no evidence that Lesson Study has turned this around. The treatment group of the first cohort is still underperforming at 45.61 points, with an absolutely negligible difference compared to the control group (at 45.49 points).

The results for the remaining secondary outcomes, the KS2 SPAG score as well as the KS2 science score, show a similar pattern. There are small differences between the outcomes of the treatment and control groups in both cohorts for GPS and in cohort 1 for KS2 science. Note that due to a policy change there exist no science scores for KS2 for the second cohort. In the right part of Table 5.2 we again present Hedges g effect sizes and 95% confidence intervals. There is no evidence for any significant performance differences between treatment and control groups, which is also indicated by p-values in the last columns well above 0.05.

Table: 5.2 Secondary outcomes

Outcome	Raw means		Raw means		n in model (interventio n; control)	Effect size	
	Intervention group	Control group	Intervention group	Control group		Hedges g (95% CI)	p-value
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
KS2 Maths							
Cohort 1	3,158	48.1	3,286	46.24	6,440	0.04	0.29
	(13)	(47.11– 49.09)	(0)	(45.28– 47.20)	(3,155– 3,285)	(-0.04– 0.12)	
Cohort 2	3,211	46.49	3,159	46.09	6,367	0.05	0.29
	(48)	(45.53– 47.46)	(31)	(45.09– 47.08)	(3,211– 3,156)	(-0.04– 0.14)	
KS2 Reading							
Cohort 1	3,226	45.61	3,323	45.49	6,545	-0.01	0.74
	(13)	(44.65– 46.56)	(0)	(44.54– 46.43)	(3,223– 3,322)	(-0.08– 0.06)	
Cohort 2	3,184	44.05	3,138	44.42	6,320	0.01	0.76
	(34)	(43.09– 45.02)	(28)	(43.45– 45.39)	(3,184– 3,136)	(-0.07– 0.09)	
KS2 GPS							
Cohort 1	3,226	47.34	3,321	46.65	6,543	0.01	0.75
	(13)	(46.39– 48.28)	(0)	(45.70– 47.59)	(3,223 - 3,320)	(-0.07– 0.10)	
Cohort 2	3,212	44.91	3,148	46.32	6,358	0	0.97
	(44)	(43.95– 45.87)	(24)	(45.33– 47.31)	(3,212– 3,146)	(-0.08– 0.08)	
KS2 Science							
Cohort 1	3,229	4.23	3,328	4.25	6,553	-0.06	0.08
	(13)	(4.20– 4.25)	(0)	(4.22– 4.27)	(3,226– 3,327)	(-0.14– 0.01)	

Section 5.2

In Section 5.1 we presented raw means of the primary and secondary outcomes for both cohorts separately. Overall, we did not see any evidence for improvements in the treatment group relative to the control group. Notably no difference in the outcomes is statistically significant at conventional levels. However, this could partly be explained by a lack of precision because we have not controlled for all observable student characteristics or unobservable school effects, only pre-treatment KS1 test scores and pair fixed effects. In this section we additionally control for student gender, FSM status when the student took KS1 examinations, and individual school effects, all of which should improve precision. Moreover, we now estimate the effects using variation within the treated schools after time by including pre-intervention outcomes for their Year 6 students and estimating a difference-in-differences randomised controlled trial design (DIDRCT). The estimates of the ITT analysis for the primary and secondary outcomes are presented in Table 5.3 and Table 5.4.

The second column of Table 5.3 reports the Beta-coefficient estimate from regressing average KS2 outcomes on an indicator variable if a school cohort was assigned to treatment or not. Cohort 1 in a school assigned to treatment will have three untreated cohorts before it also acting as control groups. Cohort 2 in a school assigned to treatment will have two untreated cohorts before it also acting as control groups. Again, we present results separately for cohorts 1 and 2. In terms of the estimates, for cohort 1 (2) we obtain -0.33 (-0.14), which are small, not significant, and even negative estimates. The corresponding p-values of the Hedges g effect size calculations (0.76 and 0.91) confirm that we cannot reject the hypothesis that the effect size is zero at any conventional level of statistical significance. This means that the ITT analysis of average KS2 outcomes confirms the previous result of section 5.1—that there is no evidence for statistically significant effects.

Table 5.3: Protocol analysis, ITT estimates, primary outcomes

Outcome	Beta (SE)	Hedges g (95% CI)	p-value
KS2 Average			
Cohort 1	-0.33	-0.01	0.76
	(1.1)	(-0.10–0.07)	
Cohort 2	-0.14	-0.01	0.91
	(1.23)	(-0.10–0.09)	

The beta coefficients represent the change in the national percentile ranking of that student due to Lesson Study. The SE figure is the robust standard error clustered at the school level.

Table 5.4 reports the corresponding set of results for the secondary outcomes, again separately for cohorts 1 and 2. For KS2 maths outcomes we estimate positive beta-coefficients of 0.35 and 0.47 percentile points. Again, these effects are small and not significant, as documented by the low values of the Hedges g effect size calculations as well as p-values close to 0.8.

Next, we reports ITT estimates for KS2 reading percentiles, separately for cohorts 1 and 2. For the first cohort we estimate a negative coefficient of 1.16 percentile points. For the second cohort we obtain a somewhat smaller negative beta-estimate of 0.69 percentile points. Notably, these (small) negative coefficients explain the overall (even smaller) negative effects of our primary outcome, namely average KS2 results as described above and shown in Table 5.3. In any case, the negative estimates of the ITT analysis for KS2 reading test scores of Table 5.4 are also not significantly different to zero at conventional levels of statistical significance. Again, this can be seen by low values of the Hedges g effect sizes as well as high p-values.

Next, Table 5.4 reports the ITT estimates for our secondary outcomes of KS2 SPAG as well as KS2 science scores. As before, we only have science outcomes for the first cohort. Again, the ITT analysis does not result in any significant patterns and fully confirms the comparison of the raw means in Section 5.1 above.

In summary, all the results from the secondary outcomes using the DIDRCT analysis confirms the previous results of section 5.1: that there is no evidence for statistically significant effects of the Lesson Study intervention.

Table 5.4: Protocol analysis, secondary outcomes; ITT estimates, secondary outcomes

Outcome	Beta (SE)	Intention to Treat (ITT) Hedges g (95% CI)	p-value
KS2 Maths			
Cohort 1	0.35	0.01	0.78
	(1.28)	(-0.08–0.10)	.
Cohort 2	0.47	0.02	0.76
	(1.53)	(-0.09–0.12)	.
KS2 Read			
Cohort 1	-1.16	-0.04	0.29
	(1.09)	(-0.12–0.04)	.
Cohort 2	-0.69	-0.03	0.53
	(1.11)	(-0.10–0.05)	.
KS2 SPAG			
Cohort 1	-0.5	-0.02	0.69
	(1.24)	(-0.11–0.07)	.
Cohort 2	-1.17	-0.04	0.38
	(1.32)	(-0.14–0.05)	.
KS2 Science			
Cohort 1	-0.02	-0.03	0.5
	(0.03)	(-0.13–0.06)	.

The beta coefficients represent the change in the national percentile ranking of that student due to Lesson Study. The SE figure is the robust standard error clustered at the school level.

Section 5.3

Tables 5.5 and 5.6 report ATE effects for specifications where we use additional information on which schools and students actually received treatment. We know that out of the 89 schools that were randomly allocated to the treatment group, only 69 schools actually implemented the full two-year Lesson Study cycle (see Figure 2). In addition, these 69 schools provided us with unique student-level UPNs that allow linking Key Stage tests for treated students directly, rather than relying on all students in the respective year groups within the treatment schools. In the previous sections 5.1 and 5.2 we compared the raw means and conducted an ITT analysis of cohorts that got the treatment assigned. However, we did not so far exploit information regarding which of the assigned schools actually followed through with the Lesson Study intervention, and which individual students were actually taught by a Lesson Study teacher. If only a small fraction actually followed through, this would be a potential explanation for the null-results that we found so far.

Moving from left to right, we first present Beta-estimates corresponding to specification 2 (page 14), where we instrument the school-cohort-level treatment with the school-cohort-level random assignment. For all outcomes, both primary as well as secondary, this analysis results in larger coefficients compared to the ITT estimates discussed above. For example, the previously estimated beta coefficient on KS2 reading percentile is now estimated at -1.76 (compared to -1.16 in the ITT). This change in size of the coefficients is expected because the ITT can be regarded as downward-biased measure of the overall effect, depending on the numbers of non-compliers, which are the schools that were assigned to the treatment group but did not actually implement Lesson Study. In the ITT we compare all schools regardless of whether they actually implemented Lesson Study, whereas the ATE analysis estimates

effects for the groups of schools that implemented Lesson Study only.³ However, the Hedges g effect size calculations as well as the p-values still do not point to differences between schools that implemented Lesson Study compared to the control group that are statistically significant at conventional levels.

Moving on to the final three columns, we present results from the student-level ATE analysis (corresponding to specifications 4 and 5). Here, we are comparing the outcomes of students for which the treatment schools sent us their individual UPN identifier in order to indicate that this student was taught by teachers who had received the Lesson Study intervention. It is worth discussing the differences that might emerge from this approach. The sampling of schools ensured that most of our schools have only one class per year group, so that effectively all students of a particular year group and cohort should have been treated if the school implemented Lesson Study. This is why, so far, the analysis has been restricted to comparing year groups across treatment and control schools. However, it is likely that schools did not implement Lesson Study for all students of a particular year group if they had two classes per year group. This would, in principle, be unobserved to us. In addition, the decision as to which students within a year group receive treatment might be non-random. The student-level ATE analysis that uses the school-level random assignment to predict treatment at the student level circumvents this problem and estimates treatment effects for treated students only.

Turning to the three last columns of Table 5.5 we can see that this analysis approach further increases the estimated effect sizes. The estimates beta-coefficients for average KS2 results for cohorts 1 and 2 are now -0.57 and -0.24. Can this be interpreted as evidence that Lesson Study actually harmed the affected students? Clearly not. These coefficients cannot be distinguished from zero at any conventional level of statistical significance. The student level-ATE analysis thus confirms the main findings in Sections 5.1 and 5.2 where we compared the raw means or conducted the ITT analysis.

Table 5.5: Protocol Analysis, ATE estimates, primary outcomes

Outcome	Effect Size					
	School ATE			Student ATE		
	Beta (SE)	Hedges g (95% CI)	p-value	Beta (SE)	Hedges g (95% CI)	p-value
KS2 Average						
Cohort 1	-0.49	-0.02	0.76	-0.57	-0.02	0.76
	(1.63)	(-0.15–0.11)	.	(1.88)	(-0.17–0.12)	.
Cohort 2	-0.22	-0.01	0.91	-0.24	-0.01	0.91
	(1.84)	(-0.15–0.13)	.	(2.08)	(-0.17–0.15)	.

The beta coefficients represent the change in the national percentile ranking of that student due to Lesson Study. The SE figure is the robust standard error clustered at the school level.

³ This interpretation of the downward bias requires the assumption that the effect size would have been the same in the treatment schools that did not implement Lesson Study.

Finally, turning our attention to the secondary outcomes (Table 5.6), the results of the student-level ATE analysis look similar, again. Notably, the previously largest negative coefficient on KS2 readings percentile scores is now increased further to -2.03 in the first cohort and -1.19 in the second cohort. Can we conclude from this that Lesson Study actually decreased reading performance? Even for cohort 1, where we estimated the larger effect, we cannot reject the null hypothesis that Lesson Study had no effect. The p-value indicates that there is roughly a 30% chance of finding this negatively estimated effect size when the true effect size equals zero. This means that none of our analysis detects statistically meaningful effects of the Lesson Study intervention on any of the primary or secondary outcomes, so far. However, finding an overall null effect could potentially hide important effects on subgroups.

Table 5.6: Protocol analysis, secondary outcomes and ATE estimates, secondary outcomes

Outcome	Effect Size					
	School ATE			Student ATE		
	Beta (SE)	Hedges g (95% CI)	p-value	Beta (SE)	Hedges g (95% CI)	p-value
KS2 Maths						
Cohort 1	0.53	0.02	0.78	0.61	0.02	0.78
	(1.9)	(-0.11–0.15)	.	(2.21)	(-0.13–0.18)	.
Cohort 2	0.71	0.03	0.76	0.8	0.03	0.76
	(2.3)	(-0.14–0.19)	.	(2.6)	(-0.15–0.21)	.
KS2 Read						
Cohort 1	-1.76	-0.06	0.29	-2.03	-0.07	0.29
	(1.67)	(-0.18–0.05)	.	(1.92)	(-0.21–0.06)	.
Cohort 2	-1.05	-0.04	0.53	-1.19	-0.04	0.53
	(1.68)	(-0.16–0.08)	.	(1.9)	(-0.18–0.09)	.
KS2 SPAG						
Cohort 1	-0.75	-0.03	0.69	-0.87	-0.03	0.69
	(1.87)	(-0.16–0.11)	.	(2.15)	(-0.19–0.12)	.
Cohort 2	-1.77	-0.07	0.37	-2.01	-0.07	0.37
	(1.96)	(-0.21–0.08)	.	(2.22)	(-0.23–0.09)	.
KS2 Science						
Cohort 1	-0.03	-0.05	0.5	-0.04	-0.06	0.5
	(0.05)	(-0.19–0.09)	.	(0.06)	(-0.22–0.11)	.

The beta coefficients represent the change in the national percentile ranking of that student due to Lesson Study. The SE figure is the robust standard error clustered at the school level.

Section 5.4: Subgroup Analysis

In this section, we present subgroup analysis in order to understand if the Lesson Study intervention had significant impact on the learning of particular groups of students. The four subgroups analysed are students who are FSM-eligible, speak English as additional language (EAL), belong to an ethnic minority, or are low achievers in terms of their KS1 outcomes.

The two tables below present the sample sizes of these subgroups as well as estimates of the corresponding ITT and the student-level ATE estimations. In terms of the results presented so far, the subgroup analysis should thus be compared to the results on the full sample presented in rows 2 to 4 of Tables 5.3 and 5.4 for the ITT, as well as the last three right-hand side columns of Tables 5.5 and 5.6 for the student-level ATE.

Turning to the subgroups, Table 5.7 shows the sample sizes and that there were 749 students of the first and 762 students of the second cohort that were eligible for free school meals. Regarding EAL status, this applied to 663 students of cohort 1 and 689 students of cohort 2, and 758 students of cohort 1 and 772 students of cohort 2 belong to an ethnic minority, where numbers in brackets refer to the second cohort. Finally, 535 students of the first cohort and 490 students of the second cohort were identified as having low KS1 result, defined as being at or below National Curriculum level 1 in either reading or maths.

Before looking at the results it is important to note that the power analysis of this trial was conducted on the full sample and not on the respective subgroups. As a result, in the subgroup analysis it will be more difficult to reject the null hypothesis of no effects for any given effect size, compared to the full sample.

Table 5.7: Number of observations by treatment status and sub group

Cohort	N in Model	Control Obs	Treated Obs	Treated Sub-Group
FSM				
Cohort 1	6,437	5,688	3,154	749
Cohort 2	6,304	5,542	3,177	762
EAL				
Cohort 1	6,437	5,774	3,154	663
Cohort 2	6,304	5,615	3,177	689
Minority				
Cohort 1	6,437	5,679	3,154	758
Cohort 2	6,304	5,532	3,177	772
Low KS1				
Cohort 1	6,437	5,902	3,154	535
Cohort 2	6,304	5,814	3,177	490

Table 5.8 displays the Hedges g effect size as well as the 95% confidence intervals and the p-value for the main effect as well as the interaction (as explained previously, the size of the differences of the effects between the groups is determined by an interaction term of assignment to treatment and the characteristics of interest). The Hedges g estimate on the interaction term is the estimated additional effect of belonging to the respective subgroup, whereas the baseline hedges-g reported in the second column refers to the effect size of the intervention for all remaining students not belonging to the respective subgroup.

Turning to the results of the subgroup analysis by free school meal status—presented in the first columns of Table 5.8—it becomes apparent that both the baseline Hedges g , as well as the interaction Hedges g , are close to zero. For example, the second row of cohort 1 shows the 95% confidence intervals which comfortably span the value of zero in each instance when moving from the ITT to the student-ATE on the right hand side. This pattern of results also holds for cohort 2, leading to the conclusion that there is no evidence that effects differ by the free school meal status of the student.

The remaining rows of Table 5.8 show the subgroup results for EAL, minority students, and low KS1 students. To cut this discussion of the table short, in no instances are there any detectable differences that are significant in a statistical sense.

Table 5.8: Main impact and additional impact on each sub group

Effect Size					Student ATE			
Cohort	Main Effect		Interaction		Main Effect		Interaction	
	Hedges-g	P-Value	Hedges-g	P-Value	Hedges-g	P-Value	Hedges-g	P-value
	(95% CI)		(95% CI)		(95% CI)		(95% CI)	
FSM								
Cohort 1	0.02	0.58	-0.01	0.73	-0.03	0.64	0.05	0.43
	(-0.06–0.10)		(-0.10–0.07)		(-0.18–0.11)		(-0.08–0.18)	
Cohort 2	0.04	0.47	-0.01	0.77	-0.02	0.79	0.05	0.43
	(-0.06–0.13)		(-0.10–0.07)		(-0.18–0.14)		(-0.07–0.18)	
EAL								
Cohort 1	0.01	0.79	-0.01	0.84	-0.01	0.86	-0.05	0.54
	(-0.07–0.10)		(-0.14–0.12)		(-0.16–0.13)		(-0.21–0.11)	
Cohort 2	0.01	0.81	-0.01	0.42	-0.03	0.72	0.14	0.36
	(-0.08–0.11)		(-0.14–0.12)		(-0.18–0.13)		(-0.16–0.43)	
Minority								
Cohort 1	0.01	0.73	0	0.94	-0.01	0.9	-0.06	0.54
	(-0.07–0.10)		(-0.11–0.12)		(-0.15–0.13)		(-0.24–0.13)	
Cohort 2	0.02	0.74	0	0.39	-0.02	0.74	0.1	0.49
	(-0.08–0.12)		(-0.11–0.12)		(-0.17–0.13)		(-0.19–0.40)	
Low KS 1								
Cohort 1	0.02	0.58	-0.01	0.73	0.02	0.58	-0.01	0.73
	(-0.07–0.10)		(-0.07–0.15)		-0.02		0.82	
Cohort 2	0.04	0.47	-0.01	0.77	0.04	0.47	-0.01	0.77
	(-0.06–0.13)		(-0.07–0.15)		-0.02		0.86	

Cost

The Lesson Study programme was implemented over two academic years (2013–2015) in 74 schools. The total cost of the implementation over this period, excluding any evaluation costs, was a little over £0.5 million. This cost information, collected from Edge Hill University on completion of the trial, can be seen in Table 6.1 (a full break down of costs can be found in Appendix C). This equates to almost £7,000 per school over the entire period, or £2,301 per year when evaluated over a three year period assuming no running costs in the third year. This equates to £53.78 per student per year, when averaged over three years. This assumes 3,167 treated students per year—the average number of students treated over both cohorts (3,157 cohort 1; 3,177 cohort 2). This is assuming all students in a school year group were treated, which may be an over-estimate or, more likely, an under-estimate as the Lesson Study model requires three teachers and there are typically fewer than three classes per year group in English schools.

Table 6.1: Expenditure for the 2013–2015 Lesson Study intervention

School Based Teacher Training		£287,820.00
VL FEES		£96,517.86
Consultant Fees		£52,436.70
Travel & Subs Staff Exp		£43,932.75
Other e.g. (Stationery, hospitality)		£11,961.13
Total Costs		£510,945.98

Another indicator of implementation costs comes from the prices for Lesson Study training published by Edge Hill University. These range from £1,800 to £2,400 per school, roughly £42–£56 per student (assuming mean student per cohort in our sample of 43). There is a discount available for groups of five or more schools.

Table 6.2: Approximate average cost per pupil over three years

	Year 1	Year 2	Year 3
Approximate average cost per pupil per year	£80.67	£80.67	£0.00

Process evaluation

Implementation

The process evaluation explored the perspectives of teachers and senior leaders on the training offered, as well as the implementation of the programme at classroom level. We carried out visits to ten schools in the South West and East of England. We selected these to ensure a mix of school size, locality, proportion of FSM pupils, and Ofsted ratings. Within each school we interviewed the lead contact with responsibility for overseeing the programme (usually a senior leader) and the teachers involved in its delivery. In most cases all such teachers were interviewed.

The process evaluation consisted of the following components:

- Evaluation of teacher preparation and training through:
 - attendance at the introductory conference in September 2013;
 - observation of teacher training in one regional centre in October 2013 and follow-up training in June 2014;
 - analysis of training evaluation data collected by the project team at Edge Hill University; and
 - analysis of data on control schools' use of peer observation and Talk for Learning approaches.
- Evaluation of implementation, fidelity, and impact through:
 - visits to ten schools in two of the three implementation regions in March 2014 to interview 19 staff and senior managers involved in the implementation of Lesson Study as team members or expert teachers;
 - follow-up interviews by telephone and email in November 2014 with five Lesson Study expert teachers in five schools, with information on progress provided by four other schools (see below);⁴ and
 - analysis of school visit data collected by the internal evaluation team at Edge Hill University.

Background

Case study and visit data found that the decision for schools to take part had been made, in most cases, by the headteacher or deputy head. Implementation was then frequently led, or overseen, by a senior leader, in many cases the deputy head. Both leaders and class teachers in the case study schools said that the impetus had come from the need to improve practice in their schools, particularly the quality of teaching and learning, and to have this reflected in KS2 results.

Previous experience of Lesson Study approaches

Schools' experiences of taking part in the project and of implementing Lesson Study related to some extent to their previous use of approaches which had ingredients of either Lesson Study or Talk for Learning. Looking first at Lesson Study itself, most of the participating schools had some experience of

⁴ We intended to carry out follow-up interviews with the lead contact (the 'expert teacher'). However, on contacting these participants in November 2014 we found a number were reluctant to be interviewed for the process evaluation for a range of reasons. These were principally staff turnover which required new Lesson Study team members to be trained. One school had decided not to continue with the intervention (the only one in the project to have done so). Rather than place additional pressure on schools at a key point in the intervention, we decided to focus efforts on collecting in-depth follow-up information from the five schools who were willing to provide this feedback.

using classroom observation, typically to assess the skills of teachers, with the focus on the teacher rather than pupils as in Lesson Study. This was usually carried out as part of appraisal or school improvement strategies but was sometimes used as part of staff development. Where used with staff development aims, some schools had practised peer observation, sharing practice between teachers, most through usually paired observations between teachers in the same or adjacent year groups. Observations were also described as fairly short, in one case only ten minutes, rather than covering a whole lesson. One case study school had brought in external input for discussions around improving practice. However, while the school had found this useful in facilitating discussion, the process as a whole was not sufficiently structured to identify areas of improvement with sufficient accuracy and detail. Teachers otherwise reported that much sharing of practice prior to Lesson Study was informal rather than structured and related to peer observation.

One case study school had called their practice 'Lesson Study' but had not included case pupils or other features of the approach. Other schools which had used sharing of practice prior to the project similarly focused on teaching, or on group learning, rather than on individual pupils. Observation had not therefore been combined with feedback. None of the participating schools had carried out pupil interviews as a means of facilitating and improving feedback, so this feature of the intervention was entirely new to all participants.

Because observation was used in contexts other than teaching and learning, and not used systematically, teachers said they knew very little about Lesson Study before they attended the training in late 2013.

Previous experience of Talk for Learning approaches

This model of Lesson Study included a focus on the quality of 'talk for learning' in guided group sessions for both English and mathematics. A number of the ten case study schools had in place an existing approach aimed at promoting talk, indicating that it is reasonably common. These approaches were described as 'Learning Partners', 'Talking Partners', and 'Talk Partners'. However, the extent to which schools used these approaches, and the areas of the curriculum in which they were practiced, varied (some schools used 'talk' approaches in literacy work, for example, but not in maths); they also varied in the extent to which they were structured and systematic. Across participating schools as a whole, data collected from school visits by Edge Hill University also found many teachers had previously used Talk for Learning approaches, including 'talk partners'. Again, Talk for literacy or writing had been more commonly practiced than Talk for Maths. Some teachers said there was 'lots of talk' in their school indicating that it was used extensively as a teaching and learning approach. However, as in the case study schools, the approach had often been unsystematic and not sustained and had rarely been formally adopted as a whole-school approach.

While Talk for Learning had not been used to its fullest extent in case study schools, those which were already experienced in such methods of improving learning found this aspect of Lesson Study easy to implement, and found they could focus more directly on the Lesson Study process.

Professional development and preparation

Following an introductory conference in September 2013 attended by senior leaders, the project team organised five days of training for teachers over the course of the intervention. These were delivered during 2013 and 2014. The content of these training sessions is shown in Figure 1 above.

Training days 1, 2 and 3 took place at the beginning of the intervention, in October and November 2013, and were designed to equip teachers with the knowledge and skills to implement Lesson Study and Talk for Learning. Day 4 was delivered in June 2014 and was designed to consolidate schools' practice and understanding of the approach, including through sharing experiences with teachers in other participating schools. Day 5 focused on using feedback within Lesson Study and took place in October 2014.

We asked teachers about their experiences of the training during our visits to schools in Spring 2014, and also analysed training evaluation data collected by the project team. We also attended training days 1, 2 and 4.

Data collected through the interviews and evaluation forms shows teachers consistently rated the training highly with teachers referring the initial training (days 1 and 2) as ‘outstanding’ or ‘high quality’, finding it well structured, interesting, and based on evidence. They liked the use of materials and group exercises. Teachers felt well prepared for implementation of Lesson Study, both in their understanding of the approach and how to use resources during lessons. While both Lesson Study and Talk for Learning aspects were found to be useful, teachers generally found Lesson Study most beneficial because, as explained earlier, they were already familiar with versions of Talk for Learning.

While the initial training was highly rated, teachers found the half day follow-up training in November (Day 3) less useful, feeling that too much was delivered in a short period of time. Teachers who had to return to their classes the same afternoon felt they had not been able to absorb the training as well as they might. The project team changed the organisation of this training session in response to this feedback. Days 4 and 5 were also rated highly by participants. At this stage in the project, teachers wanted advice with issues such as turnover of project staff (a recurring concern), and their satisfaction with the training depended to some degree on whether these were addressed rather than with core aspects of the training. These later training sessions were highly valued nonetheless, and particularly for the opportunity they gave teachers to share their experiences of Lesson Study with participants in other schools.

Practicalities of implementing Lesson Study

For schools, organisational matters included selection of teachers, including the lead ‘expert’ teacher, scheduling of Lesson Study ‘cycles’, delivery, selection of case pupils, and keeping records of sessions.

Selection of teachers

Schools used different criteria to select teachers to take part in Lesson Study. Although delivered to Year 4 and 5 pupils, the expert teacher was not required to be from those year groups. While all schools chose a subject leader in English or maths, or a teacher with specialist training in literacy or numeracy interventions as the expert teacher, selection of other members of the team sometimes included teachers whose practice was in need of improvement and schools were advised to use their own judgement on who to include. In other cases, heads chose teachers who worked well together and were therefore likely to make an effective Lesson Study team. In smaller schools with two, or even one teacher for each year group, there was little choice over the team composition and selection was not therefore strategic. It was intentional that schools were left free to choose the Lesson Study teachers themselves, so these differences do not surprise.

Teachers in the case study schools felt that the Lesson Study cycle was best delivered in a concentrated period of time, so that planning, delivery and reflection could be best co-ordinated and carried out effectively. This presented some organisational challenges, particularly in smaller schools. Schools also experienced difficulties with teacher sickness absence. In most cases problems were overcome, with schools organising cover either from existing staff or supply teachers. However, for some schools these issues led to difficulty completing all of their Lesson Study cycles.

Support and action from senior leaders, in particular headteachers, were reported as ensuring that arrangements were made and that Lesson Study teachers were given the time they needed away from their normal teaching duties. Nevertheless, teachers did feel that they would not wish to spend regular extended periods away from their own classes in the longer term, and that future application of Lesson Study would need to take this into account. They were concerned that their own class would lose out from their absence. Staff turnover, identified from all forms of data collected, was less easily addressed and resulted in the project being put on hold in some of the case study schools.

The case pupil feature

Schools were guided towards selecting case pupils from those eligible for free school meals or Pupil Premium (PP: this is extra funding offered to schools on a per-pupil basis for students from low income homes and students who have lived in local-authority care). Some schools used these more as a guide than fixed criteria and instead chose pupils who were underperforming and where the barriers to learning were not well understood. Therefore, children typically were low attainers but without Special Educational Needs. As we explain later in discussing fidelity, this criterion was seen to result in selection of pupils whose barriers were commonly found rather than pupil-specific.

Teachers generally found it easy to select pupils who they felt would be good for the project, and who would benefit from closer observation. Schools had clearly given considerable thought to which pupils would be best to select and had considered different combinations: one team had purposefully selected two children with very different barriers in order to give wider scope to include barriers for the rest of the class. The main challenge experienced was where this resulted in the selection of pupils who were not vocal. These pupils were less able to talk about their learning and contribute to the interactive parts of the intervention, particularly the pupil interviews and the feedback aspect of the intervention. However, pairing a confident and vocal pupil with a quiet one was found to work well in encouraging more active engagement.

Some teachers were concerned that pupils would find the Lesson Study set-up unusual and distracting, in particular the presence of more adults than usual in the classroom. However, this did not happen, and pupils were reported to see it as a usual activity. Some schools said this was because there were often many adults in the class so that this aspect of Lesson Study was not unusual to pupils. Others said that pupils very quickly became used to it and that it was viewed positively. Similarly, some teachers had been concerned that the case pupil feature would be noticed by other children and seen as unnatural and distracting. However, this was not found to be a problem, with children accepting that others had a particular role, or were doing special work.

Record-keeping

Teachers did not welcome the record-keeping aspect of Lesson Study largely because of the additional time commitment involved. Teachers in the case study schools, and those visited by the project team, described the demands as onerous. They saw it as a necessary and essential part of rigorous reporting, but felt that, if it continued after the project ended, a reduction in the written reporting element would be needed. A few teachers said they had not realised that they would have to write up pupil case studies after each round of Lesson Study. Knowing this at the outset would have made these first case study write-ups easier. In subsequent rounds they took notes as they went along which they found more manageable.

Fidelity

The importance of implementing the project faithfully was impressed on school leaders during the leadership training and manual; training included detailed guidance on the implementation of Lesson Study and Talk for Learning approaches. This was reinforced in all training sessions and through school visits carried out by the project team. These visits included observations of Lesson Study which were written up as visit notes by the project team and shared with NIESR as external evaluators.

Fidelity in scheduling and delivery

Visits by both the project team and by ourselves as external evaluators found schools implementing Lesson Study according to the project design. The design allowed for a degree of flexibility; for example, in the scheduling of the observed lessons, some schools had carried out the three observed lessons over consecutive days while others had done so at weekly intervals. Either approach was acceptable to the project managers, with the team requesting feedback from schools on the relative merits of either

approach. This was provided through progress reports from schools at the follow-up training days during the period of implementation.

Teachers liked the balance between prescription and freedom within Lesson Study, for example, the option to place it at the beginning of a series of lessons, or nearer to the end. The second approach was seen to allow for a concentration on the delivery and practice, rather than content. They also liked having a choice whether to interview pupils individually or in pairs.

We did come across some departure from the intended design: in one school where a member of the Lesson Study team had been off sick, the third round of lessons had been planned by one member of the team rather than the group. This departed from the design but was also found by the team to be unsatisfactory in having an adverse effect on the flow of the lesson. This was seen by the team as a learning point, in showing the value of the team approach. Another of the case study schools had put Lesson Study on hold for the autumn term of 2014 and was therefore behind with implementation, though was keeping to the format.

Fidelity in the selection of case pupils

Teachers chose case pupils according to criteria explained during the training (outlined above), but some used this more as a guide than fixed criteria. Typically, teams chose pupils who were underperforming and where the barriers to learning were not well understood. They were frequently children functioning beneath their expected level and in the lower to middle attainment grouping bands but not identified as having any Special Educational Needs. However, schools' selections did not always include pupils receiving free school meals or in the PP category as advised. Teams felt that the use of these criteria resulted in better selections than those relating to FSM or PP status. One school had used both approaches and found the FSM or PP pupils sometimes experienced atypical barriers, whereas selecting under-performers, it was felt, resulted in case pupils more representative of the educational challenges faced by the majority. Another initial concern of teachers, that the case pupil feature would be noticed and viewed as strange by pupils, was not experienced in practice. Teachers therefore felt able to deliver the project without undue concern about such issues.

A third concern among teachers related to the requirement for observing teachers not to intervene in teaching and learning, but to remain as observers. The importance of teachers observing and not intervening was emphasised particularly strongly during the training. Teachers fully understood the reason for this rule and said they followed it, though they found it difficult initially. Since we cannot tell whether teachers in schools other than those visited exercised similar restraint, we feel there may be a need for more guidance on the observing teacher role in research lessons.

Moving from fidelity to full adoption of Lesson Study

While participating schools recognised that teams should deliver Lesson Study as intended by the project team, there was evidence that some schools had become committed to using Lesson Study in a way that incorporated it into the usual practice of the school and of teachers. In other words, rather than being seen as a specific intervention with certain pedagogical approaches, methods, and practices, the principles of Lesson Study became embedded as part of a new perspective, focused on children and their learning. As one expert teacher explained:

'As you get more into Lesson Study you stop trying to fix everything at once and get much better at focusing on the key children and their particular issues with the understanding that if you can develop strategies and understanding there it will impact on others: it's about choosing the right thing for that child, the right enablers and the right concept to focus on.'

In similar vein, the expert teacher at another school explained that, nine months into the project, at the time of the third training session, her school team also 'became very focused on just what we were doing, almost forgetting the wider theory of Lesson Study'.

While schools were aware of the need to implement Lesson Study as described by the project team, a number were considering how it might be adapted for future use. Plans included reducing the number of lessons from three to two. At the same time, some of the case study schools were developing Lesson Study in ways that might be seen as enhancing rather than reducing fidelity: for example, one school was introducing 'Magpieing', where pupils are encouraged to use authors' or others' words in their own work as a way of enhancing the Talk for Learning aspect. More ways in which schools were building on Lesson Study are described below.

Outcomes

Teachers' perspectives on the value of Lesson Study

Teachers viewed Lesson Study, and the Talk for Learning approach which it incorporates, positively. They found four features of the approach useful for their own practice:

- reflecting on their teaching and learning practice;
- peer observation;
- Talk for Learning approach and resources; and
- the case pupil feature.

First, teachers welcomed the opportunity and 'space' within the timetable to reflect on their own practice, assisted by observations and feedback from teachers and pupils. They felt they learned from others' practice and appreciated contact with teachers from different year groups to their own.

Peer observation and feedback in Lesson Study was viewed as useful, particularly with its emphasis on support, rather than within performance management. The set-up of Lesson Study was seen as non-threatening and supportive. One teacher commented that the approach makes it possible to convey to an under-performing teacher what they need to do to improve in a more supportive way. Some teachers described how, as the project progressed and a relationship of trust developed, they moved from tentatively advising colleagues to more robust challenges about their teaching and learning practice.

They found the Talk for Learning aspect useful in improving the effectiveness of teaching and learning. They found the learning resources valuable, particularly those provided for the Talk for Maths element of the project.

The case pupil feature was seen as an important and distinctive element of Lesson Study, focusing attention on the learning. Unlike other observation-based approaches, the focus is on the learning and the child rather than the teacher, and was seen to lead to insights, some of which are described below. The case pupil feature was seen as valuable in allowing teachers to focus on particular barriers experienced by individual pupils.

The pupil interview was seen as a useful feature. A number of teachers commented that the experiences of children are under-appreciated, and that valuable insights can be gained from asking children about what would help their learning. One teacher commented that hearing positive feedback about a lesson from a child to a teacher can be very powerful in pointing out to fellow teachers what they need to do to improve their teaching.

Teachers also liked the fact that it does not promote a particular style, but rather a way of improving teaching and learning.

Perceived benefits of Lesson Study

Teachers identified a number of benefits of implementing Lesson Study, in particular:

- the experience of sharing practice with teacher colleagues;
- shared planning;

- identifying complementary skills;
- reflecting on pupil learning with colleagues; and
- reaching a better understanding of pupil needs.

One of the features of Lesson Study which teachers felt could improve teaching and learning was shared practice. This included having time within the school week to talk to each other about their professional practice and to observe colleagues' teaching and learning styles, approaches, and techniques. Teachers in schools visited by the project team and by NIESR said they benefitted from collaborative planning, from working alongside colleagues with different strengths, and from sharing complementary skills. Shared planning was seen as ensuring that all members of the team are involved and take responsibility for the lesson, but was also welcomed since it is an activity which is usually done alone. The shared practice aspects were seen to work best when the team worked well together.

Schools visited by the project team also talked about the benefits of having time to reflect on pupil learning and to discuss this with colleagues. Reaching a greater understanding of pupil needs through Lesson Study was seen to help with the planning of future lessons and with delivery. Teachers also talked about getting ideas from peers, including practices that were not verbally articulated and perhaps unconsciously part of another's teaching practice. This is known as tacit knowledge. Visits from the project team found teachers aware of the development of more mature relationships among staff involving the sharing and improvement of teaching and learning practice. While this can be achieved when teachers return from CPD courses, the presence in the school of a 'live' project in Lesson Study was seen as stimulating this process.

Teacher learning from Lesson Study

We have referred to a number of ways in which Lesson Study was seen by teachers as improving their classroom practice and understanding of pupils. In addition, teachers reported specific learning in relation to pupils and classroom learning but also at wider, school, level.

One teacher commented that it is possible for children to go through a whole lesson without saying a word, and that this would be missed in a big class. A teacher in a school that had already been using 'talk' approaches had found that they focused more on the content of talk when using Lesson Study, while they had previously assumed that the 'hub-bub' created by talk was productive.

Specific barriers identified through Lesson Study included poor memory and problems with comprehension. Some teachers had been made more aware of the limitations of pupils' vocabulary and how this impacted their learning. One teacher described a specific barrier when she found that, when using the prompts, pupils did not know the difference between 'describe' and 'explain'. Lack of vocabulary to explain the learning process was widely reported.

Teachers were struck by the importance to pupils of being 'right', reporting this finding at the fourth training day. Similarly, it was realised that pupils often have particular difficulty with inference within literacy, being inclined to believe there is a 'right' answer rather than a range of possibilities. The degree to which pupils wish to please their teacher also became more apparent to some teachers, particularly when conducting pupil interviews. Insights were also gained regarding the learning of more quiet or inattentive pupils who had been assumed to be not paying attention or learning. Through focusing on such children using the case pupil approach, teachers had found them to be often taking in more of the lesson than they had appreciated. This led some teachers to question the emphasis within teacher training and Ofsted criteria around looking for, and correcting, 'off-task' behaviour.

Further insight included the realisation that children choose who to listen to in class and that this impacts on their learning: some pupils contributing good responses may not be readily listened to by others. For paired learning, the impact of who children are partnered with was found to be stronger than teachers had thought. The Talk for Learning component of the intervention had revealed to some teachers the poor articulacy of some pupils, particularly in verbalising their strategies. Lack of complex vocabulary

and understanding among some pupils for whom English is an additional language was also apparent through Lesson Study and, in particular, Talk for Learning approaches. This had then been highlighted as an issue for schools to address. As one teacher commented a year into the project:

'We're asking pupils for better vocabulary but the children just don't know a wide range of vocabulary, so we've fed that back and that's having a big push in the school.'

Lesson Study also helped teachers to identify barriers to effective teaching and learning beyond the classroom: one school had improved its understanding of Ofsted's assessment that their pupils were not independent learners. This had become apparent when teachers engaged in Lesson Study observations saw that children could not explain what they had done within the lesson.

Changes in practice resulting from Lesson Study

Teachers said that Lesson Study had led to a number of changes in practice. These included changes in methods and approaches for children with identified barriers, often case pupils, in the light of insights into their learning. In particular, teachers felt they were being more effective in achieving their participation in classroom talk. In one case, a teacher described how she was now giving a reticent pupil a waiting time to contribute, rather than moving on when they did not respond. Specific practices which had been changed included greater use of enablers, such as giving pupils vocabulary lists before the lesson rather than after and allowing pupils to use icons in place of words for planning. It was seen as important to put such enablers in place for all children to avoid stigmatisation of lower achievers. Another change was in guided reading, to include greater use of inference.

A number of teachers said they had increased the use of talk in their lessons. For some, this had led to considerable change in their teaching style. They also found they had a wider repertoire of responses to use when pupils were not appearing to learn. As one expert teacher commented:

'Having been through the Lesson Study process, I find myself with a bank of ideas to try and methods to employ—more talk, less talk, different resources, colour, movement, and apply it to what we have seen with the case pupils.'

This teacher said that she and her colleagues now used 'talk' in all of their lessons. She also commented that she constantly thinks about a lesson's impact on pupils' learning:

'I evaluate every second to check that every child is having impact from the lesson and I feel confident to change it mid-lesson to use something we have found out through Lesson Study.'

Teachers said that observation continued to be used for appraisal and for school improvement but that this did not follow a Lesson Study approach. It was seen as important that it did not since Lesson Study was not aimed at producing an outstanding lesson, but focused on the barriers to learning of particular children.

Lesson Study had also resulted in other changes in school practice. For example, one school had added pupil observation to the process of recruiting new teachers, looking at the quality of applicants' observation:

'So we can see how the new teacher will be using observation to talk about the children's learning and how they can interact with a colleague to do that.'

The same school was using Lesson Study to assess the needs of pupils with special needs or disabilities (SEND).

Impact

First, it is important to note that teachers generally felt that the impact of the project resulted from the combination of Lesson Study methods and the Talk for Learning aspect, rather than either of these two separately. Talk for Learning was not seen to have an influence on its own, particularly since schools were already using a version of this already. Teachers also felt that it would be difficult to show impact

quantitatively and to link any improvement to Lesson Study, particularly for children eligible for the pupil premium who have been assisted by a range of interventions. One view was that, to have an impact, the use of Lesson Study needs to be sustained. The view was also expressed that for Lesson Study to have a significant impact on pupil performance, the whole school needs to be involved, rather than just Years 4 and 5.

Teachers believed, nonetheless, that Lesson Study was having an impact on pupil progress and was contributing to teacher development at the same time. Some teachers also said it had impacted on learning across the school. This had come about where participating teachers, and particularly senior leaders, had reported on the project at team meetings and teachers had acted on findings relating to barriers and enablers in non-Lesson Study classes.

Some teachers said they had observed a change in relation to the performance of individual case study pupils, and more widely as a result of changes they made to their teaching. These changes included improvements in vocabulary, reading comprehension, inference in enabling deeper thinking, and writing using inference. Individual pupils were reported to have become more aware of how they learn, the enablers they need, and to be more articulate about how they learn. Although a number of schools already had Talk for Learning approaches in place, the quality of talk was seen to have improved and to be taking the form of a dialogue rather than pupil statements about their learning. Pupil vocabulary was also reported to have improved.

In terms of impact on teacher practice, teachers felt that it had led to improvements, particularly among those at an early stage in their career. One expert teacher commented:

'We are all better, more reflective, teachers and we target the children positively to produce performance.'

For more experienced teachers it was seen to deepen the level of understanding of how pupils learn. It was also felt that planning within teams had improved and that team-working more generally was better as a result of Lesson Study. One expert teacher commented that the objective within the Japanese model of Lesson Study is on teacher professional development rather than pupil impact, and that for her school this was the effect they were most interested in.

A number of teachers identified the 'case pupil' aspect of Lesson Study as its most distinctive feature, bringing the focus on the impact of teaching rather than the teaching itself. Furthermore, it was seen to allow for a focus on individual children rather than groups of pupils. One expert teacher commented, a year into the project, that teachers had become better observers and more focused on children's learning.

Some teams said that, as Lesson Study had developed, they had been able to extend their focus beyond case study pupils to a wider group. This was partly because, while choosing children with specific barriers to learning, teams also found that these existed among other pupils but had not been identified.

Some teachers felt that the pupil interviews had particular impact and were one of the best features of Lesson Study. Some teams had found it more effective to interview pupils together rather than individually because this felt more natural and less intimidating. Particular benefit was identified in being able to show to pupils that their feedback had been listened to and their suggestions taken up. One limitation of the pupil interview was found in the contribution of quiet pupils. Another was the tendency among some pupils to tell teachers what they think they want to hear, to please them rather than be seen to be criticising. For this reason, one school felt that they should choose more vocal or articulate pupils in future.

The Lesson Study approach had also enabled teachers to improve pupil performance through better pairing, both of case study pupils and others. Lesson Study had highlighted the importance of balanced pairs with no dominating partner. It also became apparent to some teachers that children often have

clear ideas of who they prefer to work with and that this can be used to help ensure that pairing facilitates learning.

Finally, there was also some evidence that Lesson Study had increased schools' capacity to share good practice with other schools, with one expert teacher reporting that, a year into the project, it was working with a local secondary school interested in introducing Lesson Study as CPD for its staff.

Conclusions of the process evaluation

- The ten schools participating in the process evaluation had little previous exposure to Lesson Study and had largely experienced teacher observation in the context of school improvement or staff development. Lesson Study was seen as distinct from standard observation in its focus on the pupil and in its combination of observation and feedback.
- Teachers knew very little about Lesson Study before the intervention. This was not true of Talk for Learning, where approaches aimed at promoting talk were used to varying extents in a number of the schools. However, as with peer observation, the approach to talk had not been systematic, sustained, or used across the school.
- The project team organised five days of training during the course of the intervention. Teachers rated the training highly, finding it well structured, interesting, and based on evidence. They felt well prepared for implementation and equipped with the resources they needed.
- Implementing Lesson Study did place some demands on schools to release teachers from lessons and to allow time for planning, delivery, and reflection. Relatively few difficulties were reported, with teacher sickness and staff turnover among the few factors affecting implementation at school level.
- Teachers found the Lesson Study process—for example, selection of case pupils and observation—easier than they had anticipated. They did not welcome the record-keeping aspect of the project, though understood why it was necessary and found ways to make it easier.
- Visits by both us and the project team found schools implementing Lesson Study according to the project design and there were few concerns about fidelity. Teachers liked the balance between prescription and freedom within Lesson Study.
- There was evidence that some teachers and schools were changing their perspective by focusing more strongly on the pupil and their barriers to learning, rather than teacher delivery.
- Teachers viewed Lesson Study and Talk for Learning positively, placing particular value on opportunities to observe peers, reflect on their teaching and learning practice, use the Talk for Learning approach and resources, and focus on children through the case study pupil feature.
- Teachers valued opportunities to work alongside colleagues with different strengths and to collaborate in activities which teachers often carry out alone, for example planning and reflection.

Formative findings

A number of participating schools were planning to extend the use of Lesson Study and to continue its use beyond the life of the project. One school had already begun to use it in Key Stage 1 and was keeping to the full approach used within the EEF project. This same school was building Lesson Study into its main staff meetings and was using it as the school's main form of CPD. The school was also considering implementing it across phases, for example forming teams of teachers in Years 1, 3 and 5. One of the other case study schools had also included another teacher in the Lesson Study team to increase the school's capacity to deliver Lesson Study beyond the life of the EEF project.

However, other schools that planned to continue its use beyond the project and to roll it out to other year groups did not all intend to use the approach with the degree of intensity that the project entailed. Cost was the main reason behind any such plans, which took a number of forms. Some schools planned to use fewer rounds and to time these further apart. One school planned to continue with teams of three teachers, choosing case study pupils in line with the school's priorities but with less writing up. Another school was making similar plans. Continuing with a 'diluted' version of Lesson Study was not seen as ideal, with teachers acknowledging that Lesson Study's effectiveness stemmed, in part at least, to its intensity. Therefore, a more diluted version was seen as potentially less effective. As one expert teacher commented:

'I think probably if you diluted it you'd get diluted outcomes... I think it's about the depth and the intensity of quite a few colleagues sitting around really forensically looking at the learning with the child and that's where you get a depth of understanding.'

As with the initial decision to participate in Lesson Study, a school's decision to continue it was seen to depend on the commitment of the senior team—particularly that of the headteacher. The basis for this decision was the perceived impact and effectiveness of the approach and also its 'fit' with the school's objectives in relation to raising pupil attainment and improving the quality of teaching. Talk for Learning was seen as particularly relevant to schools' priorities, including the need to respond to Ofsted's assessments in relation to pupils' abilities to talk. The commitment of senior leaders was also seen as important for ensuring that teachers were given time and other resources to continue to deliver Lesson Study.

Although implemented by a team, it was seen as important that the project is led by one individual. One school was cementing its commitment to Lesson Study through informing the governing body about the project, its impact, and progress. Cost was the main drawback of continuing to use Lesson Study on the scale delivered within the project, or incorporating Lesson Study into a school's CPD. A particular cost was identified in the involvement of expert teachers, usually the head or deputy, and of having to use supply teachers to fill gaps. To address this, one of the case study schools was using teaching assistants to cover Lesson Study sessions delivered in addition to the project in KS1. Another factor was time spent by teachers outside of their classroom, which was seen to result in loss of continuity. For these reasons, some schools said they were considering using it on a more limited basis, for example with teachers in need of improvement in specific areas of their practice. At the same time, schools were clear that Lesson Study should not be used as part of the appraisal process itself or to judge teacher performance.

Does its evidence base influence whether schools decide to continue in its use? Teachers said that the evidence base was important in their participation in, and commitment to, the project because the investment had to be seen to be worthwhile. However, longer term use would also depend on the perceived impact on children and teachers in their school. While some of this impact might be apparent in Key Stage 2 results, some might not. Therefore, schools were interested to see whether pupils' levels of attainment improved, since this would make the case for Lesson Study more convincing to teachers. It was also said that having an evidence base, as a result of the EEF project, would encourage other schools to adopt Lesson Study.

Control group activity

In June 2016, all 181 schools involved in the project were contacted to take part in a post evaluation survey. It was agreed at the project management group meeting on 20 April 2015 that the contacting of the control schools should be left until after the final cohort of students had taken their KS2 evaluations to minimise the likelihood of influencing the control schools. This took place between 10–14 June when the evaluation team from the London School of Economics emailed all the schools with an agreed-upon survey aiming to document their professional development programmes over the previous three years. Schools that did not respond to the first email were contacted again by phone between 10–14 July alerting them to the survey and informing them that the survey would be re-sent that day, or alternatively

that it could be completed over the phone. We also told the schools throughout that we would prefer the survey to be completed by the person who was in charge of professional development in that school. The second email also contained a link to an online version of the survey, which some may have found easier to complete. No further follow-ups were conducted past 14 July, but equally, no further school completed the survey afterwards.

Of the 181 schools contacted, we received survey responses from only 32 schools, of which four (that completed the online survey) did not indicate which school they represented and so cannot be attributed to the treatment or control group. Of the remaining 28 responding schools, 17 were treated and 11 were control schools. Out of the 17 treatment schools that responded, two belong to the group that dropped out of Lesson Study right at the start and another two did not complete the second year of the intervention (but did take part in the first year). This leaves 15 replies from schools that implemented at least parts of the intervention and should thus be familiar with its ingredients.

In terms of the questions, we asked:

In the past three years has your school provided a professional development programme that included:

- teacher peer-to-peer observation and feedback;
- teacher-pupil feedback;
- talk for Learning; or
- collaborative planning?

Schools were asked if this training was self-developed or if they paid for it (which should include Lesson Study schools). The schools were also explicitly asked if they had implemented Lesson Study in the previous three years.

In terms of the results, we found that 94% of the 17 *treated* schools indicated that they had a professional development programme (PDP) that included teacher peer-to-peer observation and feedback, compared with 100% of the 11 control schools. The majority of the treated schools indicated that they had developed this programme themselves (88%) with only 35% saying that they had paid. There was similar feedback from the control schools on the questions concerning implementation of the remaining three aspects of the Lesson Study programme with 91%, 73% and 82% respectively confirming that they had such initiatives in place. These numbers were matched and slightly exceeded by the treated schools, although again, more than half of the treated schools (a similar proportion as the control schools) claimed to have developed the programmes themselves. Of course, we have to interpret these numbers with caution given the low response rate.

The goal of this intervention was to assess if this model of Lesson Study develops a superior teacher training programme compared to the status quo. Despite being limited due to the low response rate, the survey suggests that almost all primary schools have measures that resemble parts of the Lesson Study package in place already. This confirms the findings in the process evaluation discussed above, which already indicated that schools were familiar with the separate elements of Lesson Study such as teacher peer-to-peer observation and feedback. However, the treatment group received an intensive two-year Lesson Study training programme which involved a full day introductory conference as well as five separate half-day teacher trainings. In addition, the implementation of the Lesson Study treatment required documentation of the study lessons which some teachers found to be a burden. The goal of this project was to find out if this Lesson Study package makes a difference. For this, of course, we require a difference in take-up between the treatment and control schools since the effects cannot be evaluated if schools in treatment and control groups implemented Lesson Study simultaneously. To assess this possibility, we also asked schools directly about Lesson Study.

When directly asked about the implementation of Lesson Study, 14 of the 15 treated schools (that responded to the survey) stated that they had implemented Lesson Study during the previous three

years. In line with the results of the process evaluation assessment, this confirms that fidelity was high and that the treated schools really were treated.

Interestingly, four of the eleven control schools also indicated in the survey that they had implemented Lesson Study. It is possible that these schools simply became interested in Lesson Study and implemented it themselves, even though they were randomly allocated to the control group. After all, schools were randomised within a group of schools that had expressed interest in taking part in a Lesson Study RCT beforehand.

In all, the results of this survey confirm that while we were successful in ensuring a very high take-up rate among the schools of the treatment group, this happened at the potential cost of disseminating information to control schools who might otherwise not have implemented Lesson Study. Overall, however, the survey confirms that fidelity was high and that there exist large differences in the implementation of Lesson Study across treatment and control schools. All of this, however, is subject to low response rates and should be seen as indicative qualitative evidence, in addition to the school visits and interviews conducted as part of the process evaluation.

Conclusion

Key conclusions

1. The project found no evidence that this version of Lesson Study improves maths and reading attainment at KS2.
2. There is evidence that some control schools implemented similar approaches to Lesson Study, such as teacher observation. This trial might, therefore, underestimate the impact of Lesson Study when introduced in schools with no similar activity. If that is the case, the results suggest that this version of Lesson Study had no impact over and above elements of the Lesson Study approach that were already widely used.
3. Teachers felt Lesson Study was useful professional development, valued the opportunity to collaborate with colleagues in a structured way, and reported several changes to their practice as a result of the programme.
4. Schools generally implemented the programme as the developers intended. Attendance at training was high and most schools implemented one Lesson Study cycle each term.

Interpretation

This was an important trial. Of the high quality research that has been conducted on teacher development, programmes that involve teacher observations alongside effective feedback have been found to have effects (Taylor and Tyler, 2012). But there is little robust evidence of the effectiveness of such programmes. With the EEF Lesson Study evaluation we set out to change this by testing the effectiveness of teacher peer observation and feedback under experimental conditions.

We have firm confidence in the results. This was a particularly large-scale evaluation (made possible with the use of linked administrative data on national test outcomes) which included 181 schools and an overall treatment attrition rate of only 13.3% at the school level in terms of treatment and control-group allocation, and less than 1% attrition (0.002%) in terms of outcome testing. Balance was achieved on all pre-specified characteristics and the trial was powered to detect effect sizes of 0.1 with a probability of 0.8.

As well as having a well-powered trial with balance between treatment and control groups, the process evaluation reinforces our confidence in the results. Once the training had been initiated, the process evaluation found that the schools in the treatment group had a strong understanding of Lesson Study and teachers rated the training highly, finding it well structured, interesting, and based on evidence. They felt well prepared for implementation and equipped with the resources they needed. The process evaluation found relatively few difficulties with implementation, with teacher sickness and staff turnover among the few factors affecting implementation at school level.

One concern is that schools in the control group may also have been carrying out activities similar to those contained in Lesson Study. We implemented a post-evaluation survey to explore this issue directly. Of the 181 schools contacted, however, we received survey responses from only 32 schools, of which only 26 were useable, and only 11 of these from the control group (the others either dropped out of treatment before implementing Lesson Study, or did not provide details of which school they represented). This low response rate may be a result of only asking schools after the intervention had finished and after all relevant students had taken their KS2 exams. At the same time, we had to refrain from contacting schools before the end of the intervention in order to ensure no diffusion of the treatment to the control group.

While this response rate is obviously too limited to allow us to make robust inferences, the survey did suggest that almost all primary schools have measures that resemble parts of the Lesson Study

package (such as teacher peer-to-peer observation and feedback) in place already. However, the survey also confirms that the Lesson Study package is seen as something distinct; 14 out of the 15 treated schools (that responded to the survey) stated that they had implemented Lesson Study during the previous three years, confirming the case study indication that fidelity was high. However, the actual numbers of lessons taught was not recorded at the school level, so fidelity cannot be assessed beyond these qualitative indicators. Of some additional concern is that four out of the eleven responding control schools also indicated that they had implemented Lesson Study in the survey. It is possible that these schools simply became interested in Lesson Study and implemented it themselves, although they were randomly allocated to the control group. Of course, these schools did not have access to the entire training package available to the schools in the treatment group (highlighted by case study schools as impacting on professional learning for teachers). Despite its shortcomings, the survey (in conjunction with the findings of the Process Evaluation) confirms that fidelity was high and that there exist large differences in the implementation of Lesson Study across treatment and control schools.

Thus, our conclusion is that there is no effect of this Lesson Study intervention on medium-term (that is, one- and two-year) student academic outcomes for pupils on average, or among any subgroups, and we believe that these findings are robust. Note that pupils had one or two years with a Lesson-Study-trained teacher and in most cases then had one year without before their exams, and we did not evaluate any short run effects, only the effects one year after the programme had finished. This may be because the programme is not distinct enough from current practice (observations, feedback, talk for learning, and collaborative planning) in many English schools today. In contrast to the evaluation based on the external KS2 exam, the process evaluation draws a more positive picture of the evaluation, generally finding that teachers follow the protocol and find the intervention useful, particularly for supporting case pupils. We have no evidence that case pupils saw a differential effect as these pupils were not tracked individually.

Our key conclusions, therefore, are that the Lesson Study package was not successful in raising pupil performance compared to the control schools which implemented standard practice. Secondly, many schools implemented components of Lesson Study already but the additional intensity and synchronicity of the tailored two-year Lesson Study intervention has no significant impact. Thirdly, teachers viewed Lesson Study and Talk for Learning positively, placing particular value on opportunities to observe peers and reflect on their teaching and learning practice. Finally, there were no concerns about fidelity and teachers liked the balance between prescription and freedom with Lesson Study.

Limitations

As discussed, we have firm confidence in the results of this evaluation, however, as with any such evaluation, there are some caveats to bear in mind. One such caveat is that our outcome measures—KS2 reading and maths—are obtained a year after the end of implementation of Lesson Study and therefore represent a medium-term outcome. This may have reduced the possibility of finding an effect, but it is, of course, precisely where one would hope to find it. However, we cannot say what the impact might have been directly after the implementation, or indeed in the long-term, for example if there were incremental changes to teacher practice.

Additionally, our outcome measures are purely academic and therefore we cannot say whether Lesson Study may have had an impact on non-cognitive pupil outcomes such as well-being or emotional development. However, Lesson Study was never intended to affect non-cognitive pupil outcomes.

A final caveat concerns the above finding that many schools already implement parts of Lesson Study, albeit in a less structured way and without all the many elements of Lesson Study. Our research cannot say what the impact of Lesson Study would be purely on schools compared who do not carry out these individual activities. Similarly, we cannot say that any of the component parts have no impact, given that schools in the control group also perform many of them in less structured ways.

Our results are likely generalizable since they are based on a large sample of primary schools. The use of teacher observation and feedback is widespread and gaining traction, and there are many commonalities in approaches used across schools in the U.K. and internationally. The results of this research are highly relevant for schools carrying out these activities.

Future research and publications

That the Lesson Study intervention did not appear to generate any improvement in outcomes for pupils on average, or among any sub-groups, suggests that any future trials would need to consider the design of the intervention and evaluation to find significant effects. As stated above, an evaluation of Lesson Study on short-term outcomes (immediately after the programme) may yield impacts, but the results of our evaluation imply that these impacts would fade out quite quickly afterwards.

Given the suggestive evidence that many schools already implement parts of Lesson Study, one avenue for a future evaluation would be to test the impact of Lesson Study in environments where none of the individual components of Lesson Study are carried out.

References

Allen, R. (2016) 'Consistency in Key Stage 2 writing across local authorities appears to be poor', Education Data Lab, available at: <https://educationdatalab.org.U.K./2016/09/consistency-in-key-stage-2-writing-across-local-authorities-appears-to-be-poor/> [last accessed 26/5/2017].

EPPI (2003) 'The impact of collaborative CPD on classroom teaching and learning: How does collaborative Continuing Professional Development (CPD) for teachers of the 5–16 age range affect teaching and learning?', Evidence for Policy and Practice Information and Co-ordinating Centre, June 2003, Institute of Education, University of London.

Hadfield, M., Jopling, M. and Emira, M. (2011) 'An evaluation of the National Strategies Primary Leading Teacher programme', London: Department for Education.

Lewis, C. and Perry, R. (2015) 'A Randomized Trial of Lesson Study with Mathematical Resource Kits: Analysis of Impact on Teachers' Beliefs and Learning Community, in E. J. Cai and Middleton (eds), *Large-Scale Studies in Mathematics Education*, New York: Springer (pp. 133–158).

National Statistics (2016) 'Schools, pupils and their characteristics: January 2016', Department for Education, available at <https://www.gov.U.K./government/statistics/schools-pupils-and-their-characteristics-january-2016> [last accessed 19/5/2017].

Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., Martinez, A., Bloom, H. and Hill, C. (2011) 'Optimal Design Software for Multi-level and Longitudinal Research' (Version 3.01), [Software] available from www.wtgrantfoundation.org.

Rivkin, S. G., Hanushek, E. A and Kain, J. F. (2005) 'Teachers, schools and academic achievement, *Econometrica*, 73 (2), pp. 415–458.

Rockoff, J. E (2004) 'The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data', *The American Economic Review*, 94 (2).

Taylor, E. S and Tyler, J. H. (2012) 'The Effect of Evaluation on Teacher Performance', *American Economic Review*, 102 (7), pp. 3,628–51.

Appendix A: Expression of Interest form

Dear colleague,

EEF Lesson Study research project

We are delighted to offer your school the opportunity to participate in an exciting new teacher-led CPD programme.

The programme involves Lesson Study, a school led approach to CPD within which teachers work collaboratively to improve the quality of teaching and learning. Working in a small group, teachers collaborate with one another, meeting to discuss learning goals, to plan and deliver a series of actual classroom lessons (called a "research lessons"), to observe how their teaching works in practice in terms of impact on pupil learning and then to debrief on their findings in order to implement further refinements in the next research lesson.

The programme will last for two years and will be funded by the Education Endowment Foundation (EEF), which has a particular focus on raising the achievement of disadvantaged pupils. Announcing the grant, Kevan Collins, Chief Executive of the Education Endowment Foundation, said *"Lesson study is a hugely promising model of professional collaboration and, if proven to be effective, could be a method of embedding research informed teaching throughout the country."*

The programme will start in September 2013 and will run until July 2015. There is funding for 80 primary schools across 10 localities to receive locally delivered training and support, targeting Year 4 and Year 5 teachers. The pedagogic focus is on guided work embedding talk for learning in literacy and mathematics. Alongside the class teachers, schools will also be able to train a specialist such as a Curriculum or Subject Leader, MAST, AST, Leading Teacher, Numbers Count or Reading Recovery Teacher. The teachers will all be trained in both the pedagogical focus and the Lesson Study process itself.

The programme will be delivered through the Faculty of Education of Edge Hill University and led by two Expert Advisers, Di Hatchett and Gill Jordan. Participating schools will receive training and support free of charge and will also receive a grant designed to cover approximately 50% of the staff supply costs incurred. We estimate that over two years the total supply costs will be approximately £7020 for a one form entry school or £11,700 for a two -form entry school.

An important part of the study will be an independent evaluation carried out by researchers from the London School of Economics. Of the schools expressing interest, 80 will be randomly selected to participate. The impact of the programme will be assessed by comparing the Key Stage 2 attainment of pupils at Lesson Study schools with the attainment of pupils at non-participating schools. Schools will not have to provide any data for this, as the researchers will obtain information from national databases and no further information or testing will be needed. Lesson Study teachers will be asked to fill in a short survey at the end of the programme. No individual pupils, teachers, schools, or local authorities will be identified in any of the research reports.

In order to express an interest in participating please sign and return the attached proforma by Monday June 24th.

Yours sincerely,

Di Hatchett and Gill Jordan, Expert Advisers to Edge Hill University

Expression of Interest: EEF Lesson Study Project

To be returned no later than Thursday June 27th

e mail hatchett605@btinternet.com

Name of school.....

Having read the information supplied I wish to express an interest in participating in the Lesson Study programme. I understand that schools will be selected randomly and my school is not guaranteed a place. I consent to my school's data being used by for the research. I understand that the project will be evaluated using the National Pupil Database, that my school will not have to provide any data, and that no pupils or schools will be identified in the research. I also understand that, if I decide to take up an offer, my school will need to contribute approximately 50% of the supply cover costs incurred over the two years.

Signed (Head teacher)

Date

Appendix B: Memorandum of Understanding

Edge Hill University : Lesson Study Programme Supported by the Education Endowment Foundation

Partnership agreement September 2013 - July 2015

Schools will be supported to implement the Lesson Study cycle, which will be developed to focus on groups and involving teacher led guided work focused on specific aspects of English and mathematics priorities. The lesson study cycles will be undertaken by class teachers of pupils in Years 4 and 5, and the school's nominated specialist teacher.

School commitment	Edge Hill University commitment
• identify a member of the leadership team to take overall responsibility for the success of the initiative in school and to ensure the completion of all required reporting and dissemination • identify the relevant class teachers to undertake the Lesson Study cycles in English and mathematics • arrange for the involvement of a 'specialist teacher' (eg, subject leader, Reading Recovery or Numbers Count teacher, Advanced Skills teacher for English/mathematics, senior leader responsible for teaching and learning) • ensure that all the participating teachers are released in order to attend training events provided by the Edge Hill University Lesson Study programme	• provide funding to support the Lesson Study programme in the number of teachers registered • provide expert professional development for guided group work in English for participating teachers • provide expert training in Lesson Study for teachers • provide a national conference for schools

• ensure that all the participating teachers are released in order to undertake the planning, execution and debriefing involved in the Lesson Study cycles • use the core funding provided by the Education Endowment Foundation to support the implementation of the Lesson Study programme and not for any other purpose • contribute resources from within the school budget to support the release of participating teachers in order to supplement the core funding provided by the Education Endowment Foundation • support the class teachers and specialist teacher in identifying the case study pupils whose participation and progress within a guided group will form the focus of the lesson study process • support school site visits by the expert trainers assigned to the project and , where applicable, by members of the independent evaluation team • provide all data required by the programme and by the independent evaluators, including unique pupil identifiers for the pupils in Years 4 and 5 whose teachers have participated in the training programme and lessons study cycles • share teacher assessments and any other available assessments	• fully fund the costs of the training • fully fund the costs of training • provide a dedicated website for a forum for teachers to source and share practice • provide a dedicated source of communication with participating schools, a Project Leader and administrator • undertake school site visits by the project team to : - support implementation of the programme in the school; - obtain a qualitative view of the impact of the programme on the school; - gather case study material and evidence of the impact of learning and pupil progress • work with the independent evaluators to collect data and report on the impact of the programme • liaise with relevant personnel in the relevant agencies , to keep them informed of any issues arising and to invite them to contribute • convene an expert advisory group to provide a challenge to the project national
---	--

Edge Hill University : Lesson Study Programme 2013-15

Supported by the Education Endowment Foundation

Logos to be added EHU, LS, EEF

1. Having considered the respective commitments set out in the partnership agreement provided, I confirm the acceptance of the place offered on the EEF funded lesson Study programme 2013-15.

2. Having read the detailed information provided about the CPD sessions to be attended and the requirements undertaken in school, I wish to register the following teachers for participation (**minimum** per school: one Year 4 'expert'; **maximum** per school : two year 4 teachers and two year 5 teachers plus one 'expert'). If you are unable to please indicate roles.

Teacher 1 Role..... Name.....

Teacher 2 Role..... Name.....

Teacher 3 Role..... Name.....

Teacher 4 Role..... Name.....

Teacher 5 Role..... Name.....

Signature	Job title	
	Headteacher	
	Project Leader	Edge H

Instructions for signing and returning to be added. I suggest return date of July 9th

Appendix C: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. Cost ratings are awarded as follows:

Cost rating	Description
£ £ £ £ £	<i>Very low:</i> less than £80 per pupil per year.
£ £ £ £ £	<i>Low:</i> up to about £200 per pupil per year.
£ £ £ £ £	<i>Moderate:</i> up to about £700 per pupil per year.
£ £ £ £ £	<i>High:</i> up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high:</i> over £1,200 per pupil per year.

The Lesson Study programme was implemented over two academic years (2013/14-2014/15) in 74 schools. The total cost of the implementation over this period, excluding any evaluation costs was £510,945.98. This cost information was collected from Edge Hill University who implemented the Lesson Study programme upon completion of the trial (See table below) This equates to £6,904.68 per school over the entire period or ££2,301.56per year when evaluated over a three year period assuming no running costs in the third year. This equates to £53.78 per student per year, when averaged over three years. This assumes 3167 treated students per year which is the average amount of students treated over both cohorts (3157 cohort 1, 3177 cohort 2). This is assuming all students in a school year group were treated, which may be an over or more likely an under-estimate as the Lesson Study model requires three teachers and there are typically fewer than three classes per year group in English schools.

Breakdown of costings

EXPENDITURE:	
SCHOOLBASED TEACHER TRAINING	£287,820.00
FACILITIES HIRE	£52,436.70
CONSULTANTS FEES	£43,932.75
TRAVEL & SUBS. STAFF EXP.	£18,277.54
VL FEES	£96,517.86
VL GROSS	£4,562.64
TEMPORARY STAFF	£5,706.95
RECHARGE-PHOTOCOPYING, EHC	£669.99
RECHARGES-CATERING COSTS ETC	£554.15
RECHARGE-STATIONERY-SUBSIDS	£93.00
HOSPITALITY	£117.56
SCHOOL PRACTICE, STAFF TRAVEL	£18.80
COMPUTING CONSUMABLES	£60.00
STATIONERY	£178.04
TOTAL EXPENDITURE FROM EEF FUNDS	£510,945.98

Appendix D: Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition*					
5 	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%	5 				5 
4 	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%			Adjustment for Balance [] Adjustment for threats to internal validity []		
3 	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%					
2 	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%					
1 	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	51-50%					
0 	No comparator	MDES > 0.6	>50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 5 padlocks less than 1% attrition
- **Reason for adjustment for balance** (if made):
- **Reason for adjustment for threats to validity** (if made):
- **Final padlock score:** initial score adjusted for balance and internal validity = 5 padlocks

*Attrition should be measured at the pupil level, even for cluster trials.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

OGL This information is licensed under the Open Government Licence v3.0. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/>

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at www.educationendowmentfoundation.org.uk



Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk